Handbook of the Logic of

Argument and Inference

The Turn Towards the Practical

edited by Dov M. Gabbay, Ralph H. Johnson,

Hans Jürgen Ohlbach and John Woods

CONTENTS

PREFACE

The *Handbook of the Logic of Argument and Inference* is an authoritative reference work in a single volume, designed for the attention of senior undergraduates, graduate students and researchers in all the leading research areas concerned with the logic of practical argument and inference. Here is how the work came about.

One August in the nineties, researchers from disciplines as disparate as computer science, logic (formal and informal), philosophy of language, psychology and argumentation theory assembled in Dagstühl, Germany, for a week-long stay. The purpose of the gathering was to determine whether the will and the intellectual wherewithal existed to bring to the problem of practical reasoning some degree of interdisciplinary insight. The main emphasis of the workshop was on the sort of theory it would take to qualify as a successful logic for practical reasoning.

After several days of stimulating papers, and much fruitful discussion, it was determined that the relevant research communities would be well-served by a *Handbook*, and editors were appointed and given the task of designing and executing such a work. Anyone who has worked in an interdisciplinary arrangement will appreciate the challenges posed in crossing subject-lines. One of the heavier costs, although not the only one, is the demands such arrangements make on time. Another is the effort required to evade the inevitable charge of dilettantism pressed against researchers in their home discipline when they venture beyond.

Over time the Editors gravitated to a certain view of what the handbook should do.

1. It should reflect the fact that the interdisciplinary approach to practical reasoning is not yet an accomplished fact, that it is currently an enterprise that we meet with *in media res*.

2. It should reflect the fact that there is much common agreement that since classical logic was not designed as a logic of practical reasoning, it is not an adequate general theory of it.

3. Even so, it should reflect the fact that as matters now stand, the research programmes in practical reasoning constitute a rich and complex pluralism.

Accordingly, the *Handbook* has been given the following design. After an introductory chapter, the role of standard logics is surveyed in two chapters. These chapters can serve as a mini-course for interested readers, in deductive and inductive logic, or as a refresher. Then follow two chapters

of criticism; one we call the internal critique and the other the empirical critique. The first deals with objections to standard logics (as theories of argument and inference) arising from the research programme in philosophical logic. The second canvasses criticisms arising from work in cognitive and experimental psychology. The next five chapters deal with developments in dialogue logic, interrogative logic, informal logic, probability logic and artificial intelligence. The last chapter surveys formal approaches to practical reasoning and anticipates possible future developments. Taken as a whole the *Handbook* is a single-volume indication of the present state of the logic of argument and inference at its conceptual and theoretical best. Future editions will periodically incorporate significant new developments.

The Editors wish to thank most warmly the following persons and agencies for valuable assistance:

- Dov Gabbay and Hans Jürgen Ohlbach for organising the Dagstühl workshop.

- Jane Spurr in London and Dawn Collins, Stephen Friesen and David Graham in Lethbridge, for technical and production support.

- The Dean of Arts and Science at the University of Lethbridge, Professor Bhagwan Dua, for financial support of this project.

- In addition to funding agencies mentioned by the authors we would like to thank The Physical Sciences and Engineering Research Council of the United Kingdom and the Social Sciences and Humanities Research Council of Canada.

- The many excellent colleagues who read chapters in draft and offered helpful suggestions: David Corfield, Luíz Leite, Donald Gillies, Erik Krabbe, João Leite, and Halina Przymusinska.

- Arjen Sevenster, our Publisher at Elsevier Science BV for wise and unstinting support.

The Editors
London, May 2001.

LIST OF AUTHORS

Else M. Barth
Praediniussingel 10/8,
9711 AG Groningen,
The Netherlands.
henkelse@xs4all.nl

J. Anthony Blair
Department of History, Philosophy Program,
University of Windsor,
Windsor, ON, N9B 3P4, Canada.
TBlair@uwindsor.ca

Dov M. Gabbay
Department of Computer Science,
King's College, Strand,
London WC2R 2LS, UK.
dg@dcs.kcl.ac.uk

Ilpo Halonen
Department of Philosophy,
P.O. Box 24, 00014 University of Helsinki,
Finland.
ilpo.halonen@helsinki.fi

Gilbert Harman
Department of Philosophy,
Princeton University,
Princeton, NJ 08544-1006, USA.
harman@princeton.edu

Jaakko Hintikka
Department of Philosophy,
745 Commonwealth Avenue,
Boston University,
Boston, MA 02215, USA.
hintikka@bu.edu

Ralph H. Johnson
Department of History, Philosophy Program,
University of Windsor,
Windsor, ON, N9B 3P4, Canada.
johnsoa@uwindsor.ca

Arto Mutanen
Department of Philosophy,
P.O. Box 24, 00014 University of Helsinki,
Finland.
arto.mutanen@helsinki.fi

Hans Jürgen Ohlbach
Ludwig-Maximilians-Universitaet Muenchen,
Institut fuer Informatik,
Oettingenstr. 67,
D-80538 Muenchen, Germany
ohlbach@informatik.uni-muenchen.de

Luis Moniz Pereira
Departamento de Informática
Faculdade de Ciências e Tecnologia ,
Universidade Nova de Lisboa, 2829–556 Caparica,
Portugal.
lmp@di.fct.unl.pt

David N. Perkins
Project Zero, Harvard Graduate School of Education,
124 Mount Auburn Street,
Cambridge, MA 02138,
USA.
e-mail address needed

Jon Williamson
Department of Philosophy,
King's College London,
Strand, London WC2R 2LS
jon.williamson@kcl.ac.uk

John Woods
The Abductive Systems Group,
University of Lethbridge,
4401 University Drive ,
Lethbridge, AB, T1K 3M4, Canada.
woods@uleth.ca

JOHN WOODS, RALPH H. JOHNSON, DOV M. GABBAY
AND HANS JÜRGEN OHLBACH

LOGIC AND THE PRACTICAL TURN

1 PURPOSE OF THIS HANDBOOK

The chapters of this volume exhibit a kind of internal dialectic. Chapters 2 and 3 are pivotal. They record standard positions in logic from which present-day theories of argument and inference are in process of retreat. Chapters 4 to 10 are in various ways expressions of *protest* against the standard positions.[1] The drift of these chapters is toward the practical, toward psychologically real models of human interaction in real time. Accordingly, the volume is "book-ended" by chapters on this theme. Chapter 1 establishes a contrast with the standard approaches of the coming two chapters; and chapter 11 sets the stage for still future developments, especially as regards the construction of formal models. It is one of the burdens of the *Handbook* to reveal something of current thinking about the inadequacy of standard logics. In taking this critical stance, its authors defer to no one in their admiration for the myriad accomplishments of standard logics of deduction and induction. The focus of the criticism is a highly specific one. Such logics don't do well as theories of real-life argument and reasoning. In so saying, the burden shifts. What logics will accomplish this task better?

Quine has famously quipped, "Logic is an old subject and since 1879 it has been a great one." No one will deny the momentous developments that logic has undergone in the past century and a quarter; but there are those, among whom we count ourselves, who doubt that logic's greatness is so recent and comparatively brief a thing. For all their mathematical splendour, the modern logics of deduction, induction — and to a lesser extent abduction — stand out for their fidelity, advertent or otherwise, to Peirce's insistence that logic has nothing to do with how we think, and that a logic of practical reasoning — of vital affairs, as Peirce says — is impossible. Peirce's theoretical austerity has not, even so, discouraged logicians from thinking that the rules of logic serve as *ideal norms* for human reasoning. Either way, whether Peirce's or that of ideal norms, actual thinking and reasoning, thinking and reasoning *on the hoof* as we shall say, is significantly underdetermined by the providence of standard logic.

It is worth emphasizing that the distance between theory and practice did not inhibit legions of logicians in the first two-thirds of the century just past, and legions still today, from asserting salient connections with real-life

[1] Thus, the 'Ambitious claim' is false. See Perkins, chapter 5 below.

reasoning and arguing. This was done for the most part with acknowledgement of the desirability to expose the structure of practical reasoning as it occurs in real-life, in exchanges, both spontaneous and deliberative, under actual conditions.

The mathematical turn in logic has never been without its critics, needless to say, whether Sidgwick [1910], Schiller [1929] or Dewey [1938]. But by and large these were not mainstream views. The emphasis on the practical could not, however, be denied indefinitely; and in the 1970s there arose criticisms of mathematical logic, both from sciences external to logic — chiefly computer science, psychology and linguistics — and, more strikingly perhaps, from within the philosophical mainstream of logic itself. New disciplinary arrangements arose out of these dissatisfactions with standard logics as theories of inference and argument, notably argumentation theory itself — an amalgam of logic, probability theory, fallacy theory, discourse analysis, conversational analysis, cognitive science, computer science, forensic science and rhetoric.

There is as much dissimilarity, and more, in these various approaches to reasoning and arguing. But virtually without qualification, they can be seen as converging on a common theme. It is that the peculiarities of practical reasoning call out for, and deserve, theoretical attention, if not from standard logic, then from a logic suitably revised for the task; and even if thus revised, not by logic alone, but by a practical logic in concert with various purpose-built sister disciplines that have arisen in that past generation.

This marks what we may fairly call a *turn toward the practical*. Every chapter of this *Handbook*, beyond the surveys of Chapters 2 and 3, illustrates this turn in important ways, whether in the precincts of informal logic, dialectical logic, interrogative logic, probability logic and computer logic; or argumentation theory and cognitive science. Although the turn toward the practical is evident in these developments, it is also clear that we meet with the turning *in medias res*. There is nothing in this fruitful motley that stands out as a mature and settled theory, and there is nothing discernible in them that qualifies as a developed *practical logic*. The turn toward the practical is, therefore, an historical event in progress. It represents a development in logic which is as significant as the mathematical turn of a century and a third ago; for it is a transition that promises a return of logic to its historic objective as a principled account of how real people transact their inferential and argumentational agendas on the hoof, whether in the *agora*, or on the shop floor, in the corporate head office or Cabinet room. It is the purpose of this *Handbook* to chart the turn toward the practical, such as we have seen of it up to now, and to give some indication of the targets on which what remains of the turning may be expected to lock.

Practical reasoning has been a recognized species of thinking since antiquity. Given its importance and its ubiquity, practical reasoning is a natural

target for the theorist. We can trace the study of practical reasoning back to Aristotle (384–322 BC) who, like many who followed, appears to differentiate between practical reasoning (aimed at some end or action) and theoretical reasoning (aimed at a cognitive good of some sort — a theory, hypothesis, explanation etc.) We shall later have more to say about this distinction.

Early theorists saw a structural similarity between reasoning (of a given kind) and arguments (of that same kind). Here again Aristotle was the first to press the similarity in a theoretically explicit way. Given the close connection between reasoning, argument and deduction, it is natural that the dominant view of these matters has been that a good theory of deductive argument is likewise a good theory of deductive inference. The closeness of the connection has been asserted from ancient times to the present, with recurring disagreements as to which relation of the similarity is primary. It remains a contentious issue, as we see in this passage from Dummett, (in which "assertion" can without loss be replaced by "argument", and "judgment" by "inference"):

> We have opposed throughout the view of assertion [argument] as the expression of an interior act of judgment [inference]; judgment [inference], rather, is the interiorization of the external act of assertion [argument]. (Dummett [1973, 262])

This view looms large as we move to the developments in logic in the 19th and 20th century. Our own view lies closer to Davidson's, to the effect that

> neither language nor thinking can be fully explained in terms of the other, and neither has conceptual priority. The two are, indeed, linked in the sense that each requires the other in order to be understood, but the linkage is not so complete that either suffices, even when reasonably re-inforced, to explicate the other. (Davidson [1984, 156])

The Editors share the conviction that the *Handbook* should be an authoritative reference resource for any researcher or senior student active in any of the fields that presently converge on the issue of practical reasoning. Every chapter is written in the recognition that for some readers it must serve as an introduction to the disciplines and research problems with which the chapter deals. It is also important that these chapters be of interest to specialists. Chapters 2 and 3 are a case in point. They are a survey of standards approaches to deduction and induction, and as such must cover the basics. At the same time, these chapters are written with a respect for the rigour appropriate to their disciplinary standards, and they develop material which some specialists will not have encountered before, or not have encountered in this form.

Since Aristotle, it has been customary to suppose that there is a close relationship between logic and reasoning. This accounts for the assumption, made by many theorists, that *logic is a theory of reasoning*, a vexed assumption, as we have seen. By "logic," the most recent tradition tends to think in terms of modern formal deductive logic, or *FDL* for short. However, we note in passing that Aristotle did not understand the logic of deduction in the way that classical logicians of the present day do. For classical logicians (and lots of non-classical ones as well), logic is a theory of validity, a theory of the consequence relation. For Aristotle, logic is *not* a theory of validity or logical consequence. It is a theory of — to coin a term — *syllogisity*, or syllogistic consequence. Syllogisms are valid arguments which satisfy additional (and powerful) conditions. One is that there can be no syllogism with redundant premisses. Another is that no premiss can be repeated as a conclusion. A third bans multiple conclusions. The import of these constraints can be seen in the kind of logic that they produce. It is striking that the first logic in the Western tradition should have been, or should greatly have approximated to, a linear (hence relevant), intuitionistic, nonmonotonic, paraconsistent logic. This matters for the claim that the logic of deduction is also a realistic theory of deductive reasoning. For, again, deduction here is not the deduction of the logics of the modern mainstream, but rather the deduction of something manifestly more psychologically real than it.

For good or ill, in these past two and a half millennia it has been natural to assume that logic is the theory of reasoning. From time to time, logic has been specifically adapted, as in the case of deontic logic, to take account of practical reasoning. Notwithstanding some aggressive skepticism over the ages, logic's hegemony in these matters persisted until the latter part of the twentieth century. Even to this day, many logicians introduce their subject to first-year students as an account of deductive *reasoning*, and nearly everyone persists in the habit of calling the transformation rules of deductive logic "rules of inference." If ever a case could have been made for supposing that the rules of the logic of deduction are indeed rules of inference available to real-life reasoners in real-life situations, it would have been a conception of deduction that satisfied the more psychologically real constraints on syllogisms. Since 1879 which marks the publication of Frege's *Begriffsschrift*, the classical logic of deduction honours none of these constraints, making the claim that logic is a theory of reasoning a good deal less plausible.

Nor should we omit to remark that even syllogistic logic has been called into question over the years for various kinds of inadequacy. In the period of European modernity (roughly from 1600 AD to the end of World War II), the then standard logic intermittently attracted some harsh assessment. Writers as different from one another as Bacon (Bacon [1960]), the Port Royal Logicians (Arnauld and Nicole [1996]), Locke (Locke [1975]), Whately (Whately [1848]), Mill (Mill [1974] and Alfred Sidgwick (Sidg-

wick [1910]), shared a general kind of criticism of School Logic, the form in which syllogistic logic was then expressed. They saw the logic of syllogism as psychologically unreal. It was not well-suited, and did not apply in a natural way, to the reasoning of everyday concerns. And it was no good, they thought, as a logic of discovery. Against this we find the powerful dissent of Peirce [1992], in which the very idea of a logic of everyday affairs is dismissed as absurd. Mill takes a more moderate tone. The correct logic of scientific reasoning will also be the correct logic of business and practical affairs. Such a logic is not however the logic of syllogisms. But this did not keep Mill from writing as follows:

> I know nothing, in my own education, to which I think myself more indebted for whatever capacity of thinking I have attained.
> (Mill [1974, Chapter 1, section 12])

Locke's diffidence was even more emphatic, placing at risk his own anti-syllogistic complaints:

> And I readily own that all right reasoning may be reduced to his [i.e., Aristotle's] forms of the syllogisms
> (Locke [1975, Chapter 17, Book Four]).

All the same, Locke adds, syllogisms are not the only nor the best ways of reasoning, for leading those into truth who are willing to find it. The Port Royal logicians, Arnauld and Nicole, proved the stouter critics. Syllogistic logic may have its place in a more comprehensive logic of science, but a logic of science could never suffice for practical affairs. Sidgwick in his turn acknowledged that there were two main conceptions of how logic should be done. Writing after Frege (see Sidgwick [1910]), it is perhaps not surprising that he should have recognized the use of mathematical methods and mathematical notations in logic. On the other hand, when it comes to systematizing the reasoning of everyday common sense, a practical logic is to be preferred, or a logic involving the direct scrutiny of arguments from daily life, to which could be adapted some of the findings of the empirical and forensic sciences.

It would be wrong to suggest that these critics did not have their day. The Port Royal Logique has had an influence on university curricula in France and beyond which has persisted to this present day. The influence of Mill's *A System of Logic* was not as long-lived but was substantial even so. It was adopted as a textbook first at Oxford and subsequently at Cambridge.

2 LOGICAL LATITUDE

Important as these criticisms of FDL were, they were insufficiently robust to place at any risk the hegemony of formal symbolic logic from Frege onward. It has proved next to impossible to discourage in any operationally

significant way the habit in modern logicians of seeing their enterprise as one that delivers a principled account of reasoning. Given the steadfastness of this *idée fixe* it is reasonable to suppose that, for this question of the state of logic at the beginning of the new millennium, the following are the historically dominant developments, on top of the founding of logic itself by Aristotle:

- The establishment of the logic of propositions by Megarian and Stoic logicians
- The revival of systems of formal dialectic by mediaeval logicians
- The creation of the probability calculus by Laplace, Pascal and other seventeenth century thinkers
- Modern mathematical logic, independently co-established by Frege and Peirce, and the subject of a very large number of important ramifications in the century following 1879

- Cognitive psychology, from the 1970s onwards
- Computer science and AI, developed during this same period
- Informal logic, Critical Thinking and Argumentation Theory, also developed in this period.

There is a natural break in this list after the entry for mathematical logic. (It is marked by the line.) Dominantly a logic of properties of propositional structures, such as the consequence relation, mainstream logic after Frege had few of the user-friendly constraints of the logic of syllogisms, as we have said; and went on unimpeded to significant intellectual achievements in its four principal domains of model theory, proof theory, recursion theory and set theory. (See here Gabbay and Woods [2000a] and [2000c]) Non-mainstream logics were another matter. Relevant logic took on, unawares, one of the syllogistic constraints, as did intuitionistic logic a different syllogistic constraint. Modal logics appropriated their operators from Aristotle's modal logic, as well as the variations of mediaeval writers. And so on. It is not entirely surprising that a Fregean logic, together with the mathematical theory of chance, should have become the targets of criticisms from every theoretical perspective on our list which occurs *below* that line.

Part of the task of the *Handbook* is to tell the story of this conflict between what is above the line and what is below, with special attention to contexts of practical reasoning. To this end, it will be a substantial convenience to confine our multiple narratives to approximately the last one hundred and twenty-five years, with a special emphasis on developments over the last four

decades. It is true, of course, that some of what cognitive scientists find fault with in mainstream logic, and some of what AI theorists find it important to play up, was prefigured in flourishes of criticism at various times in the past 2500 years. As we have said, much of what contemporary informal logicians and argumentation theorists see fit to criticize in mainstream logic can be found in writers such as Arnauld, Locke, Mill and Sidgwick. In a way, these historical truths don't matter, for the critical developments presently on view were made for the most part from the 1970s onwards, and largely in ignorance of, or anyhow without reference to, the earlier intimations of them that dot our intellectual history.

Given these comments about this history of logic, how shall we understand logic for our purposes here? It is a heftily ambiguous word. This is not a wholly decisive consideration, but it does suggest what we think of as wise counsel. Better to give the word free reign (provided that the requisite distinctions are honoured), than to privilege a narrower use which, as is known in advance, lots of people won't honour anyway.

There are three elements that may be thought of as central to the idea of logic and we shall say something about each in turn. They are (1) *normativity* — that is, it is thought that logic provides norms for reasoning rather than descriptions of how people reason; (2) *systematicity* — which in its purest form is the idea that logic develops decision procedures or implementation protocols for the questions it faces; (3) *foundationality* — that is the idea that logic is methodologically, or epistemically prior, to other disciplines (some would say all other disciplines). We shall briefly discuss each of these.

(1) *Normativity.* As will be seen in Chapter 2's discussion of modern mathematical logic, a standard criticism is that its models aren't psychologically real. For one reason or another these are models which no actual human reasoner — not even the most able among us — could conform his inferential behaviour to. Most theorists concede this point; but some defenders of modern logic argue that even so, its legitimacy is preserved under the appropriate idealizations.

The distinction between what holds of actual reasoners and what holds of ideal reasoners also bears on a further contrast of fundamental importance to the theorist of practical reasoning. (It also bears on the question of how free a range to give to the word "logic".) On a widely received view, a theory of reasoning is an authoritative normative account if, and only if, its claims are true in the appropriate ideal model. For example, if a rule of inference is honoured in the model, this shows the rule to be one which actual reasoners ought to comply with, even if they do not do so in fact. On the other hand, a theory of reasoning is an authoritative descriptive theory if its claims are verified by the behaviour of actual human reasoners, irrespective of whether contrary claims are upheld in the appropriate ideal model. This leaves the theorist with one good (and difficult) question. Are

there non-question-begging ways of specifying ideal models, fidelity to which qualifies a theory as genuinely normative? For the purpose of this present chapter, we leave this as an open question.

(2) *Systematicity.* In taking a systematic approach to practical reasoning, we do not presume strict computability in the sense of computable functions and the like. We welcome computability where there is evidence of its presence. But what we more generally presuppose is susceptibility of reasoners to problem-solving capabilities subject in principle to decomposable description.

Decidability places highly on the contemporary logician's list of metamathematical virtues. This is odd since, apart from it's monadic sublogic, decidability is not a virtue that even classical first order logic can lay claim to. Of course, the classical logic of sentences is decidable, but standard systems of relevant sentence logic — system R, for example — are not. (See Dunn [1984]) Although some systematicizations are decision procedures, the more basic idea is that of a set of *general rules* which regulate the performance of a given task — the construction of a proof, the making of an inference, the reaching of a decision, and so on. Rules here are understood in the manner of procedures, and procedures in turn are understood as ways of getting certain things done, in some cases tacitly. Related to this is the idea that a logic — a logic of practical reasoning, for example — gives a good account of itself to the extent to which it is able to define procedures for the competent production of practical reasoning (or whatever else the target of the account in question might be). If the theorist in question is also a computer scientist, the present idea he will see as mother's milk, viz., as instructions embedded in his software.

This is a basic idea for us, too. But it is necessary to sound a caveat. If we think of certain kinds of reasoning — of inference, say — as a matter of belief-revision or belief-update, then it matters a lot what the theorist takes beliefs to be. If he thinks that he can make do with the idea that belief is (sociolinguistic) assent to a sentence or some such thing, then the present notion of systematicization is well-defined for belief-revisions and belief-updates. For on this account, rules are things which the inferer, i.e., the language-user, can obey or disobey at will. But if beliefs are taken to be psychological states, they are not easily seen as being commanded by the will, as assent surely is. This being so, algorithms for inference are not convincingly representable as rules in the sense of the lines just above, but rather as *state conditions*, as we might call them. State conditions are causal processes which operate beyond the direct reach of the subject's will, and largely automatically. (See Harman, Chapter 4 below, Woods [2001f, Chapter 4] and Gabbay and Woods [2002].)

It is perhaps nothing to be alarmed about if for theoretical reasons the concept of rule were stipulated to cover both kinds of mechanisms, leaving it to context to determine whether rules in the more common sense meaning

are intended, or whether it is state conditions on belief-revision and belief-update that the theorist has in mind. We are ourselves are inclined toward the latter view, in as much as rules in both sense share a common input-output structure. So for us systematicization is essentially a matter of performance rules in this broad sense, rules which we might, for reasons that are now evident, describe as *virtual*.

(3) *Foundationalism*. The foundationalism we assume for these various approaches to practical reasoning is not the big architectonic of, say, logicism. Logicism is a failed programme in the foundations of mathematics. It was an attempt to show that all of standard mathematics reduces without relevant loss to quantification theory plus axiomatic set theory. So conceived of, logic plus set theory was both foundational for and exhaustive of mathematics. We ourselves are not so minded. But we take it to be an assumption worth exploring that there exists a *base logic*, or a *protologic* for practical reasoning, or failing that, a family of logics — a partially ordered set — partially ordered under the relation of *being more fundamental to* an algorithmic account of practical reasoning. And failing *this*, we would settle for some reasonable approximation of such a poset.

Here, too, where evidence of undergirding structures is available in ways that offer promise of axiomatization, we do not disdain it. But our assumption is that our various disciplines can rightly aspire to the exposure of features connectable by partial orderings, from the more fundamental to the less, and so on. There is the related question of whether the disciplines that presumptively fulfil our runner-up subtitle have a suitably general name. On one way of looking at things, they all qualify for the loosely rendered name of logic. This is not to overlook those legions of researchers for whom the word "logic" has been reserved, by God or some suitably positioned Platonic overseer, for a theoretical working up of the consequence relation and kindred additional properties, but nothing else, in a formal language, or what Church calls a *logistic system* (Church [1956, 48]).

As we said at the beginning of this chapter, the sequence of papers in this volume in some sense reflects the journey we have just been describing. As we have said, we will be focussing our efforts mainly on the last 125 years which have been marked by the emergence and domination of *FDL*. Chapters 2 and 3 will survey that dominance in the theory of argument and inference. These chapters set a standard of sorts. The chapters to follow chronicle the breakdown of that standard. It is a breakdown that is still in progress.

As we have indicated, the limitations of *FDL* as anything like a theory of reasoning emerged in criticisms that will be discussed here. In Chapter 4, Harman presents what we have called the *internal critique*. Its internality is a function of the fact that it is a criticism which arises within the philosophy of logic itself. In Chapter 5, Perkins discusses what we call the *empirical critique* — the ways in which the story told by FDL turns out to

be controverted by empirical research.

These limitations of modern logic (and others) have inspired a series of reform or breakaway logics, and these are discussed in the remaining chapters. In Chapter 6, Barth gives a thorough treatment of what is called *dialogue logic*; in Chapter 7, Hintikka, Halonen and Mutanen give an account of *interrogative logic*; in Chapter 8, Johnson and Blair develop their treatment of *informal logic*. Williamson's Chapter 9 chronicles dissensions which have emerged in the *logic of probability*; and Chapter 10 is given over to Pereira's account of salient developments in *computer science*. Finally, in Chapter 11, Gabbay and Woods set out their foundational proposals for a practical logic. In all of these, it will be evident that the authors are moved by what they take as limitations of modern logic. For Harman the problem is intractability; for Barth it is deductivism; for Hintikka, the failure to focus on strategic rules; for informal logic, over-concern with formalism; for computer logic, the over-abstractness of *FDL*; for probability logic, intractability is again a problem; and for dynamic logicians, it is *FDL*'s indifference to temporality.

3 PRACTICAL REASONING

In one sense of the term, all reasoning is practical.[2] All reasoning terminates in an answer to a question, a solution to a problem, a conclusion from some data, or a decision to postpone the quest until further facts are known; even aborted reasoning produces a kind of termination. Against this is the point that a great deal of human reasoning is automatic, subconscious and pre-linguistic, concerning which the careless postulation of goals could be needlessly anthropomorphic.

A case in point: You step off the curb, intent on crossing to the other side of the street. A horn blares to your left. You step back onto the curb. A good thing too, you think, as a pick-up truck whizzes by. True enough, you evaded the truck. But how helpful is it to claim on this account your possession and activation of a *truck-evasion strategy*? (But see 1.4 below.)

Ordinary usage, even ordinary philosophical usage, gives little direct guidance for fixing the sense of practical reasoning. It is an expression layered with multiple meanings and suggestive of contrasts, among which are these:

ordinary, common *versus* esoteric, specialized
prudential *versus* alethic
moral *versus* factual
informal *versus* formal
precise *versus* fuzzy

[2]There is a philosophical tradition in which a practical reason is reason for an action that involves bodily behaviour. Needless to say, not all reasoning is practical in this sense. We ourselves are disposed to think that practical reasoning in this sense hardly carves out a natural kind, so to speak. (See here, e.g., Velleman [2000]).

conclusion is an action *versus* conclusion is a proposition
premiss is an action *versus* premiss is a proposition
goal-directed, purposive *versus* context-free
applied *versus* theoretical
concrete *versus* abstract
tolerant of incommensurabilities *versus* not

To these we add a further contrast, to which we think it prudent to take particular note of. It is the contract of

practical *versus* strict

We illustrate with an example. In the game of (ice) hockey, a hat trick is achieved by a player scoring three consecutive goals against the opposition. "Consecutive" here means "without any goal being scored between the first and the third of this triple by any of the hat tricker's team-mates". This is what a hat trick is *strictly speaking*. But *in practice*, or for *all practical purposes* (including the triggering of bonus clauses in a player's contract), a hat trick is just three goals in a game by one and the same player, never mind whether he scores them consecutively in our present sense of that term. So conceived of, practicality is resemblance enough to the real thing to be considered the real thing. At the same time, it carries the suggestion of something (admissibly) subpar. We will briefly revisit this contrast in section 5 below.

In the chapters to follow, most of the issues implied in this variegated usage will be taken up and examined. Our earlier latitudinarian policy toward "logic" serves as a helpful guide here. For here too there are good reasons not to strive for a single canonical meaning of "practical reasoning" and, correspondingly, for a privileged theoretical approach to it. The fact is that in its present state, the business of practical reasoning is business on several different fronts. It is an attractive explosion of conceptual and theoretical possibilities. We would do well to accept this diversity, as much anyhow as we can justify, rather than attempt to highjack the concept of practical reasoning under threat of a charge of self-servingness.

4 A PROTOLOGIC: IDEALIZED LANGUAGE

In Chapter 11, Gabbay and Woods present a way of approaching a theory of practical reasoning — a *protologic*, as it were. For the present, we offer the merest sketch of this approach, as it will help to clarify how we see the coming together of logic and practical reasoning — and hence provide further orientation to the chapters in this volume. We do so without hegemonic intent and without having forgotten the proposed flexibility with which we closed the preceding section. Ours is but one way in which to get the busi-

ness of a theory of practical reasoning done in an intellectually satisfying way.

We saw at the beginning of the present chapter the makings of a dispute between those who see all reasoning as practical and those who claim a principled division to the contrary. Seen the first way, reasoning that strikes us as automatic, subconscious and extralinguistic still qualifies as practical provided that we admit tacit goals, virtual aims, and the like. Not everyone will relish the prospect of endless wrangles about the existence of such purported entities. Against this, people of a more sanguine mien may find such disputes refreshing, and in any case possessed of a more fundamental attractiveness than might initially have been supposed. Our own view is that any way of drawing the practical-non-practical distinction in a plausible way will be subject to some sort of actual-virtual distinction with respect to whatever qualifies the reasoning as practical. It is better to try to draw a serviceable distinction between practical reasoning and, whatever we take its contrast to be, in ways that don't require us to give a detailed account the actual-virtual distinction.

We join a large consensus in thinking of an episode of practical reasoning as directed to or productive of an action in real time. How nearly this forecloses on the practicality of purely private, purely mental reasoning is in large measure a matter of how well disposed we are towards psychological doctrines of mental states and mental acts. Those who distrust folk psychology (e.g., Churchland [1981]) or doubt the bona fides of introspection (e.g., Dennett [1996]) are free to fix the limits of practical reasoning as they see fit. Our own inclination is (again) one of tentative latitudinarianism.

Our approach emphasizes action and real time. It acknowledges the attractions of dialogical factors in developing an adequate theory of practical reasoning. It recognizes that it lies in what Malraux called the human condition that so much of what human beings do has a social dimension.[3] There may well be a place for reasoning in the head, but from infancy onward a great deal of it occurs interactively with others. (Privacy has to be learned.) A natural and near-ubiquitous medium of these interactions is the dialogue (from the Greek *dia* for "across or between", and *logos* for "talking"). A conspicuously important subcase is the dialectical form of a dialogue — a form of enquiry or attack shaped by question and answer.[4]

Even solo reasoning is transacted with resources the reasoner could not have had except for prior intercourse with his fellows. Solo reasoning can be seen as a form of enquiry, a kind of interrogation of Nature herself. (Hintikka [1989], Hintikka and Bachman [1991], and Hintikka, Halonen and Mutanen, Chapter 7, this volume). But it is not our view that interior

[3]We could as easily here mention such widely diverse philosophers as Aristotle and Peirce.

[4]Contemporary usage shows scant regard for the difference between dialogues and dialectical exchanges. We ourselves see little virtue in such indifference.

monologues are truncated conversations with an alter ego, transacted in the language of thought, or mentalese. Our interest in interactive reasoning is that it is reasoning-on-the-hoof, relatively easy to identify and to try to make something of theoretically. Interactive reasoning instantiates the old saw, 'transparency in publicity'. Neither do we suppose that a theory of interactive reasoning will adapt in every way to what goes on strictly in the head. But it would be folly to suppose that a good account of social reasoning has nothing to teach us about our solo efforts, if any. So we emphasize practical interactive reasoning because we think that the relevant regularities are often easier to spot.

It also bears on our question that dialogue logic is inherently a practical logic in our present sense. Dialogues engender commitments, which impose obligations and responsibilities that parties are expected to meet by taking the appropriate action at the appropriate time. It does not diminish the point at hand to note that often, indeed typically, those actions are of an expressly dialogical or linguistic character, as with assent and dissent.

Everyone agrees that humans operate under resource limitations. Unlike the Enlightenment view of the individual, that human beings are not contingently related to these traits. We think of the human individual as intrinsically social and intrinsically resource-challenged. This being so, human actions, no matter the circumstances, reflect the kind of being the human individual is. It also means that a practical logic must be a, so to speak, a social logic. And although dialogue-settings are a recurring and typical context for an individual's actions, a practical logic need not, as we have said, be confined to a dialogue logic, although it should contain dialogue logic as a sublogic.

It is well to note at this juncture two generic ways of conceiving of logic. One involves thinking of logic as a set of properties defined for *idealized languages*. The other sees logic as a set of properties ascribed to *ideal agents*. We close the present section with some remarks about the first conception. The section to follow is reserved for the second conception.

For ease of exposition, we will speak of linguistic and agency conceptions of logic, respectively. It is easy to see that these two are prefigured in the original ambiguity of syllogism, an ambiguity that has endured ever since. This is the ambiguity reflected in the contrast between logic as a theory of argument and logic as a theory of reasoning. From the very beginning logicians have honoured a conception of argument in which arguments are linguistic structures. Not every logically significant conception of argument is linguistic in this way, but in virtually every major development of the subject, logic has found room for such a conception.

If, on the other hand, logic is taken as a theory of reasoning, a theory of inference, then it is natural to think of it as giving a principled account of what the *thinking agent* does, and also what happens to him under certain circumstances. It is not as natural to simulate this idea of logical agency

by constructions on linguistic structures, although this has not stopped logicians over the centuries from trying to do this very thing.

In its linguistic conception, it is not difficult to see that our previously noted triple of concepts — normativity, systematicity and foundationalism — can be engaged rather naturally. Of the three, systematicity secures the readiest purchase. If the heart and soul of systematicity is input-output mechanisms under the drive of rules, language is tailor-made for rules. Under suitably regimented conditions, a language's entire syntax can be generated by rules, and its semantics by functional connections between linguistic structures and set theoretic structures.

The present authors have expressed their doubts about the normative *bona fides* of logic, but for those who are differently minded, normativity can be thought of in part as language idealized or regimented — as canonical notation as Quine calls it — and as rules which are truth-preserving.

Foundationalism comes most naturally in the distinction between axioms and (derived) theorems, but is also caught in the distinction between primitive and derived rules, as well as in the distinction between a logic as a set of sentences and a superlogic, i.e., a nonconservative extension of it.

A further attraction of the linguistic conception of logic is its openness to metatheoretic manipulation. Sets of sentences are more or less public entities, out there in eminent domain, so to speak. On the present view, logic is linguistics on purpose; it chases truth up the tree of grammar, in Quine's witty epigrams. Just as logic drapes its target properties on syntactic structure, so too does metalogic exploit these same grammatical forms.

The peculiarities of what we have been calling our protologic are also easy to specify, with designated lexical items for times and action, and rules of derivation purpose-built for sentences contriving such expressions.

5 A PROTOLOGIC: IDEALIZED AGENCY

On the agency view, logic is a theory of reasoning, a theory of what thinkers do and have happen to them. Correspondingly, a practical logic is a theory of what practical agents think and reflect upon, cogitate over and decide, decide and act. If the linguistic conception makes it centrally important for the logician to say, with care, what sort of thing a language is, the agency view makes it necessary to say, with care, what sort of thing a practical agent is. With this in mind, our protologic must make room not only for times and actions, but also for the agents who perform those actions at those times.

We think of practical agency as a *hierarchy* of goal-directed, resource-bound entities of various types. At the bottom of this hierarchy are individual human beings with minimal efficient access to institutionalized data

bases. Next up are individual human beings who operate in institutional environments — in colleges or government departments, for example, which themselves are kinds of agents. Then, too there are teams of such people. Further up are disciplines and other corporate entities such as the American Space Program or Soviet Science in the 1970s. The hierarchy proceeds thus from the concrete to the comparatively abstract, with abstract structures being aggregations of entities lower down. Interesting as this metaphysical fact might be, it is not the dominant organizing principle of the hierarchy. The organizing principle is economic. Entities further up the hierarchy command resources, more and better, than those below are capable of.

So conceived, the hierarchy is a poset of objects partially ordered by the relation of *commanding greater resources than*.[5]

Every agency in this hierarchy $\langle C, A\rangle$ involves, whether by aggregation or supervenience or in some other way, the individual agent. Such agents are thus basic to any logic of agency, and it is to them that we shall concentrate our attention in the present section.

Like all agents in the hierarchy, the individual is a performer of actions in real time. And nearly everything an individual is faced with doing, or is trying to do, can be done at the wrong time. It can be done at a time so wrong as to court equivalence with not doing it at all, or doing some opposite thing. It is not enough that an agent does the right thing, i.e., performs the right action-types. It is very often essential that the right thing be done at the right time. As we look upwards at the agency-hierarchy, we see a diminishing susceptibility to exigent timeliness. No one doubts that NASA had a real deadline to meet in the sixties, culminating in the moon shot. It might have been that the moon program would have been cancelled had that deadline not been met. Even so, individuals are exposed to myriad serious dangers, many of them mortal, that nothing up above will ever know on this scale; and essential to averting such dangers is doing what is required on time.

The dominant requirement of timeliness bears directly on a further constraint on individual agency. Individuals wholly fail the economist's conceit of perfect information. Agents such as these must deal with the nuisance not only of less than complete information, but with data-bases that are by turns inconsistent, uncertain, and loosely defined. To these are added the difficulties of real-time computation, limited storage capacity and less than optimal mechanisms for information-retrieval, as well as problems posed by bias and other kinds of psychological affect. (See Harman, chapter four below.)

[5]We note in passing the difference of our hierarchical model from Harry Frankfurt's hierarchical model of autonomous action. On this latter conception, the behaviour that an agent makes happen in the fullest sense of that expression is that which is motivated by a desire which the agent desires to have. See Frankfurt [1988, 58–68]. But cf. Bratman [1999, 185–206].

The two great scarcities that the individual must cope with are time and information. It is precisely these that institutional agents command more of, and very often vastly more of. With few (largely artificial) exceptions, the individual agent is a satisficer rather than an *optimizer* a fact reflected in our distinction between the practical and the strict, captured by the example of the hat trick in section 3 above. For the most part, even seeking to be an optimizer would be tactically maladroit, if not actually harmful. The fact of the robust, continuing presence of human agents on this Earth amply attests to their effective and efficient command of scarce resources. It is a fact in which is evident the human capacity to compensate for scarcities of time and information.

We postulate that the individual agent embodies a **scare-resource compensation strategy**. Here in rough outline, and no particular order, are the compensation-factors that strike the present authors as particularly important.

- Human beings are natural **hasty generalizers**. It was a wise J.S. Mill who observed (Mill [1974]) that the routines of induction are not within the group of individuals, but rather are better-suited to the resource capacities of institutions. The received wisdom has it that hasty generalization is a fallacy, a sampling error of one sort or another. The received wisdom may be right, but if it is, individual human agency is fallacy-ridden in degrees that would startle even the traditional fallacy-theorist.[6] Bearing on this question in ways that suggest an answer different from the traditional one is the fact that the individual's hasty generalizations seem not to have served his cognitive and practical agendas all that badly. Upon reflection, in the actual cases in which a disposition towards hasty generalization plays itself out, the generalizations are approximately accurate, rather than fallacious errors, and the decisions taken on their basis are approximately sound, rather than exercises in ineptitude. Not only is the individual agent a hasty generalizer, he is a hasty generalizer who tends to get things right.

- How is it possible that there be a range of cases in which projections from samples are so nearly right while at the same time qualifying as travesties of what the logic of induction requires? The empirical record amply attests to a human being's capacity for pre-inductive generalization and projection. It would appear that exercise of this capacity involves at least these following factors, some of them structural, some of them contextual. The *pre-inductive* generalizer does not generalize to universally quantified conditional propositions. Rather

[6] On what we are calling the traditional account of fallacies, hasty generalization is always an error. For contrary view see Woods [2001e].

he generalizes to *generic* propositions. There is a world of difference between "For all x, if x is a tiger then x is four-legged" and "Tigers are four-legged". The former is falsified by the truth of any negative instance, whereas the latter holds true even in the light of numerous negative instances of certain kinds. We could characterize this difference by saying that universally quantified conditional statements are highly *fragile*, whereas generic statements are robust.

The robustness of what the pre-inductive generalizer generalizes to serves the generalizer's interests in other ways, two of which are particularly important. One is that the individual agent is a *fallibilist* in (virtually) everything he thinks and does. The other is that the individual agent has the superficially opposite trait of very *high levels of accuracy* in what he thinks and does when operating at the level ordained for him by the hierarchy of agency. Generalizing to generic statements is a way of having your cake and eating it too. It is a way of being right even in the face of true exceptions. It is a way of being right and mistaken concurrently.

Generalizing in this way also works a substantial economy into the individual's cognitive effort. It comes from the smallness of its samples and the robustness of its generalizations. Generic inference is inference from small samples under conditions that would make it a fatally stricken induction. We see in this the idea of the affordable mistake. Generic inference is not truth-preserving. One can be wrong about whether Pussy the tiger is four-legged even though one is right in holding that tigers are four-legged. Affordable mistakes are like small infections that help train up the immune system. Just as an infant's summer sniffles is an affordable (in fact, necessary) infection, so too are the small errors of the cognitive agent which provide him evolving guidance as to the freedom and looseness with which to indulge his predilection for comparatively effortless generalizations. Baby's summer cold loops back benignly in the discouragement of more serious illness. Affordable mistakes loop back benignly in the discouragement of serious error. We can now see that the old saw of *learning from our mistakes* has a realistic motivation. We do not learn from mistakes that kill us.

- What is it about such samples that sets them up for successful generic inference? It would appear that the record of generic inference is at its best when samples, small as unit sets though they may be, are samples of **natural kinds**. There has been a good deal of philosophical controversy about whether natural kinds actually exist; about whether the putative difference between natural kinds and conventional kinds turns on a principled distinction. Certainly there is nothing like a settled consensus as to how the distinction should be applied. Perhaps

this tells against our here using the concept in any theory-laden way, but it leaves it open that we introduce it as a term from unanalyzed common sense. Even so, we should not disdain this literature from psychology and computer science in which concepts resembling that of natural kinds seem to be doing useful work, concepts such as those of *frame* (Minsky [1975]), *prototype* (Smith and Medin [1981]), and *exemplar* (Rosch [1978]).

Philosophical particularities aside, the empirical record testifies to our capacity for classifying sensory stimuli in ways that reflect similarities and differences that strike us as inhering things as they really are. There is ample evidence to suggest that our classifications originate with primitive devices of *type-recognition* together with the mechanisms of fight and flight. It is significant that some of our most successful and most primitive inferences involve the recognition of something as dangerous. Generic inference is part and parcel of such strategies. Just as our capacity for recognizing natural kinds exceeds the comparatively narrow range of immediately dangerous kinds, so too does our capacity for generic inference exceed the reach of fight-flight recognition triggers. But whether in fight-flight contexts or beyond, natural kinds and generic inference are a natural pair. It is an arrangement again favouring the economic — a compensation strategy for the scarcity of time and information — but not noticeably at the cost of error. If generic inferences from natural kind samples are not *quite* right, at least they don't kill us. They don't even keep us from prospering.

- The fallibilism of generic inference is also evident in its relation to **defaults**. A default is something taken as holding, taken to be true, in the absence of indications to the contrary. Most of what passes for common knowledge is stocked with defaults, and generic inferences in turn are inferences to defaults. Default reasoning is inherently conservative and inherently defeasible. Defeasibility is the cognitive price one pays for conservatism. And the great appeal of conservatism is also economic. Conservatism is populated with defaults in the form "X is what people have thought up to now, and still do". Conservatism is a method of default-collection. It bids us to avoid the cost of fresh thinking, and to make do with what others have thought before us (and, experienced and remembered, too).

  Conservatism places a premium on what is already well-received. On the face of it, conservatism is the *ad populum* fallacy in endemic form. Here, too, we might grant the received wisdom (and note the large irony), conceding that individual agents are notorious fallacy-mongers on a scale not dreamed of even by the traditional fallacy theorist. But as we said in our examination of a similar indictment involving the

charge of hasty generalization, there are factors which seem to cut across so harsh a condemnation. One is that we are, by and large, enormously well-served by the trust we place in the testimony of others. This needs to be understood. The full account, even if we could furnish it, is beyond the scope of this chapter, but certain features stand out, and should be mentioned. It is well to note in passing that the phenomenon we are presently reflecting on is *popular belief*, not popular *taste*. Popular taste is often taken to be vulgar, sentimental, prurient, or substandard in other ways. It would be a mistake to try to reconcile the idea of popular belief to this model. Popular beliefs are what Aristotle called *endoxa*. They are "reputable opinions", the opinions of everyone or of the many or of the wise. The mere fact of popular opinions triggers an abduction problem. What best explains that p is a proposition believed by everyone? An answer, which certainly can be criticized in respect of certain particular details, but which cannot convincingly be set up for general condemnation is that p's universal acceptance is best explained by supposing that p is true or that a belief in p is accurate. What is loosely called common knowledge is an individual's (or an institution's or a society's) inventory of endoxa. What is especially striking about common knowledge is that it is acquired by an individual with little or no demonstrative effort on his own part, and with attendant economies of proportional yield.

- It is evident therefore that individual agents depend for what they think and for how they act on the sayso of others, on the more or less uncritical and unreflective testimony of people who by and large are strangers. Here is yet another respect in which the conduct of human agents would seem to fall foul of the received opinion of fallacy theorists (let us not forget that the endoxa of the wise are not guaranteed to be true!). For it would appear that individual agents are programmed to commit and implement the program on a large scale, the *ad verecundium* fallacy. But as before, the actual record of thoughts and actions produced by such dependancies is rather good; most of what we think in such ways is not especially inaccurate and, in any case, not inaccurate enough to have made a mess of the quotidian lives of human individuals. We may suppose, therefore, that the traditional fallacies of hasty generalization, *ad populum* and *ad verecundium* are hardly fallacious as such (e.g., when considered as an individual's strategies or components of strategies for practical action), but are fallacies only under certain conditions. We shall return to this point below.

It has long been known that human life is dominantly social, and that individual agents find cooperation to be almost as natural as breathing. The routines of cooperation transmit to an individual nearly all

of the community's common knowledge that he will ever possess.

Even though the complete story has yet to be told, cooperation has received the attention of attractive and insightful theories (e.g. Axelrod [1984], Coady [1992]).

There is a natural and intuitive contrast between accepting something on the sayso of others and working it out for oneself. Cross-cutting this same distinction is the further contrast between accepting something without direct evidence, or any degree of verification or demonstrative effort on the accepter's part, and accepting something only after having made or considered a case for it. The two distinctions are not equivalent, but they come together overlappingly in ways that produce for individual agents substantial further economies.

Perhaps this is the point at which to emphasize that in our conception the individual is not the artefact of the same name, championed by European thinkers of the seventeenth and eighteenth centuries. We demur from the notion (the decidedly odd notion, as we see it) that an individual's social relationships are merely contingent to his rationality. On the contrary, an individual's cognitive and decisional competence is in significant part constituted by his social relationships. If this is right, it will matter for what we take a logic of individual cognitive and decisional agency to be. We will have more to say on this later, but will note in passing the prima facie attractions of a dialogue logic, as a formalized description of the individual agent.

Such additional economies are the output of two regularities evident in the social intercourse of agents. One has been dubbed the reason rule:

> *Reason Rule*: One party's expressed beliefs and wants are a *prima facie* reason for another party to come to have those beliefs and wants and, thereby, for those beliefs and wants to structure the range of appropriate utterances that party can contribute to the conversation. If a speaker expresses belief X, and the hearer neither believes nor disbelieves X, then the speaker's expressed belief in X is reason for the hearer to believe X and to make his or her contributions conform to that belief.
>
> (Jacobs and Jackson [1983, 57], Jackson [1996, 103]).

The reason rule reports an empirical regularity in communities of real-life discussants. Where the rule states that a person's acceptance of a proposition is reason for a second party to accept it, it is clear that "reason" means "is taken as reason" by the second party. Thus a descriptively adequate theory will observe the Jacob-Jackson regularities as a matter of empirical fact. This leaves the question of whether

anything good can be said for these regularities from a normative perspective. If normativity is understood as a matter of instrumental value, it would appear that the reason rule can claim some degree of normative legitimacy. Not only does it produce substantial economies of time and information, it seems in general not to overwhelm agents with massive error or inducements to do silly or destructive things. The reason rule describes a default. Like all defaults, it is defeasible. Like most defaults, it is a conserver of scarce resources. And like many defaults, it seems to do comparatively little cognitive and decisional harm.

There is a corollary to the reason rule. We call it the *ad ignorantiam* rule:

> *Ad Ignorantiam Rule*: Human agents tend to accept without challenge the utterances and arguments of others except where they know or think they know or suspect that something is amiss.

Here, too, a good part of what motivates the *ad ignorantiam* rule in human affairs is economic. People don't have time to mount challenges every time someone says something or forwards a conclusion without reasons that are transparent to the addressee. Even when reasons are given, social psychologists have discovered that addressees tend not to scrutinize these reasons before accepting the conclusions they are said to endorse. Addressees tend to do one or other of two different things before weighing up proffered reasons. They tend to accept this other party's conclusions if it is something that strikes them as *plausible*. They also tend to accept the other party's conclusion if it seems to them that this is a conclusion which is within that party's competence to make — that is, if he is seen as being in a position to know what he is talking about, or if he is taken to possess the requisite expertise or authority. (See, e.g., Petty and Cacioppo [1986], Eagly and Chaiken [1993], Petty, Cacioppo and Goldman [1981], Axsom, Yates and Chaiken [1987], O'Keefe [1990], and the classic paper on the atmosphere effect, Woodworth and Sells [1935]. But see also Jacobs, Allen, Jackson and Petrel [1985]). We see, once again, the sheer ubiquity of what traditionalists would call — overhastily in our view — the *ad verecundiam* fallacy.

We see the individual agent as a processor of information on the basis of which, among other things, he thinks and acts. Researchers interested in the behaviour of information-processors tend to suppose that thinking and deliberate action are modes of consciousness. Studies in information theory suggest a different view. Consciousness has a narrow bandwidth. It processes information very slowly. The rate of

processing from the five senses combined — the sensorsium, as the Mediaevals used to say — is in the neighbourhood of 11 million bits per second. For any of those seconds, something fewer than 40 bits make their way into consciousness. Consciousness therefore is highly entropic, a thermodynamically costly state for a human system to be in. At any given time there is an extraordinary quantity of information processed by the human system, which consciousness cannot gain access to. Equally, the bandwidth of language is far narrower than the bandwidth of sensation. A great deal of what we know — most in fact — we aren't able to tell one another. Our sociolinguistic intercourse is a series of exchanges whose bandwidth is 16 bits per second (Zimmermann [1989]).

Conscious experience is dominantly linear. Human beings are notoriously ill-adept at being in multiples of conscious states at once. And time flows. Taken together these facts loosely amount to an operational definition of the linearity of consciousness. Linearity plays a role in the cognitive economy that tight money plays in the real economy. It slows things down and it simplifies them. Linearity is a suppressor of complexity; and reductions in complexity coincide with reductions in information.[7]

Psychological studies indicate that most of our waking actions are unattended by and unshaped by mental states. This mindlessness of ordinary waking human behaviour is a kind of coping. Consider a case in which we are watching a short-order cook working at full blast at a New York midday. It is easy to see his behaviour as connectionist and mindless, as behaviour reflecting repetories of different skills which he draws upon concurrently and distributively, and without a jot of reflection when things are going well.

Here is a view that carries interesting consequences for the analysis of propositional assent — indeed for conversation generally. On the received view, when someone asserts, e.g., that the cat is on the mat, he in effect reports a current mental state (a belief, say), an object of my present attention. His assertion is sincere if the belief actually exists, and true if the belief is also true. Conversation more generally still is a sequence of exchanges of transparent propositional contents *modulo* the usual speech acts.

If these psychological studies are right, the received view is wrong. Conversation would just be linguistic coping. If so, the individual discussants are less often in a state of belief than many theorists suppose;

[7]We note in passing that the sheer paucity of information possessed by human consciousness at any given time contrasts with environments known to be fuzzy. Fuzziness, unlike probability, is unchanged by arbitrarily large increases in information.

and when someone is telling us about the amenities of, say, Amsterdam, though he tells us the truth, he is not transmitting any current mental state and he is not inducing new mental states in us, unless perhaps what he tells us is surprising. When we stop and think — when we put a temporary (and expensive) halt to coping — we find that in what we do in the world we are infrequently the owner of mental states, infrequently the possessor of beliefs. It is a respectable way of being mindless.

It is now evident that we must amend the claim that individual agents suffer from a scarcity of information. In so doing, however we are able to lend appropriate emphasis to what remains true about that proposition. In pre- or sub-conscious states, human systems are awash in information. Consciousness serves as an aggressive suppressor of information, preserving radically small percentages of amounts available pre-consciously. To the extent that some of an individual's thinking and decision-making are subconscious, it is necessary to postulate devices that avoid the distortion, indeed the collapse, of information overload. Even at the conscious level, it is apparent that various constraints are at work to inhibit or prevent informational surfeit. The conscious human thinker and actor cannot have, and could not handle if he did have, information that significantly exceeded the limitations we have been discussing. This makes the *economic* aspect of an agent's conscious thought and action an *ecosystemic* matter as well. Human beings make do with slight information because information is all the conscious individual can have.

Human agents make do with scarce information and scarce time. They do so in ways that make it apparent that in the general case they are disposed to settle for *comparative* accuracy and *comparative* sensibleness of action. These are not the ways of error-avoidance. They are the ways of fallibilism. Error-avoidance strategies cost time and information, except where they are trivial. The actual strategies of individual agents cannot afford the costs and, in consequence, are risky. As we now see, the propensity for risk-taking is a structural feature of consciousness itself. It might strike us initially that our fidelity to the reason rule convicts us of gullibility and that our fidelity to the *ad ignorantiam* rule shows us to be lazily irrational. These criticisms are misconceived. The reason rule and the *ad ignorantiam* rule are strategies for minimizing information over-load, as is our disposition to generalize hastily.

Consciousness makes for informational niggardliness. This matters for computer simulations of human reasoning. That is, it matters that there is no way presently or foreseeably available of simulating or mechanizing consciousness. Institutional agencies do not possess

consciousness in anything like the sense we have been discussing. This makes it explicable that computer simulations of human thinking fit institutional thinking better than that of an individual. This is not to say that nothing is known of how to proceed with the mechanization of an individual's conscious thinking. We know, for example, that the simulation cannot process information in quantities significantly larger than those we have been discussing here.

Consciousness is a controversial matter in contemporary cognitive science. It is widely accepted that information carries negative entropy. Against this is the claim that the concept of information is used in ways that confuse the technical and common sense meanings of that word, and that talk of information's negative entropy overlooks the fact that the systems to which thermodynamic principles apply with greatest sure-footedness are *closed*, and that human agents are hardly *that*.

The complaint against the over-liberal use of the concept of information, in which even physics is an information system (Wolfram [1984]), is that it makes it impossible to explain the distinction between energy-to-energy transductions and energy-to-information transformations. Also singled out for criticism is the related view that consciousness arises from neural processes. We ourselves are not insensitive to such issues. They are in their various ways manifestations of the classical mind-body problem. We have no solution to the mind-body problem, but there is no disgrace in that. The mind-machine problem resembles the vexations of mind-body, both as to difficulty and to type. We have no solution to the mind-machine difficulty. There is no disgrace in that either.

- For individual agents it is a default of central importance that most of what they experience, most of what is offered them for acceptance or action stands in no need of scrutiny. Information-theoretic investigations take this point a step further in the suggestion that consciousness itself is a response to something disturbing or at least peculiar enough to be an interruption, a demand — so to speak — to pay attention. If this is right, consciousness is an aberrant state, the exception rather than the rule, and the same is true both for case-making and for the consideration and evaluation of cases. This affects practical reasoning in an especially interesting way, for it squares with a practical reasoning problem requires a *trigger* and that a trigger is an event or state of affairs or scrap of information which stands out in some way, which demands attention and calls for an explanation.

- Most of the information processed by an individual agent he will not attend to, and even if it is the object of his consciousness he will attend to in as little detail as the exigensies of his situation allow. Arguing is a statistically nonstandard kind of practice for human agents, but even when engaged in it is characterized by incompletions and short-cuts that qualify for the name of enthymeme. The same is true of reasoning, of trying to get to the bottom of things. In the general case, the individual reasoner will deploy the fewest resources that produce a result which satisfies him. Here is further evidence that individuals display a form of rationality sometimes called "minimal", and well-discussed in Cherniak [1986].[8] In addition to features already discussed in this chapter, the minimal rationalist is, when he reasons at all, a non-monotonic reasoner and in ways that are mainly automatic, the successful manger of belief-sets and commitment-sets that are routinely inconsistent. Much of what makes for the inconsistency of belief-sets comes from the inconsistency of deep memory storage and further aspects of inconsistent belief-sets flow from the inefficiencies of memory retrieval.

The structure of minimal rationality shows the individual agent to be the organic realization of a nonmonotonic paraconsistent base logic, features which our protologic must take care to embed. There is little to suggest that the strategies endorsed by classical logic and most going nonstandard logics form more than a very small part of the individual agent's repertoire of cognitive and coping skills. If it is true that individuals are in matters of non-demonstrative import *pre*-inductive rather than inductive agents, the same would also appear to be the case as regards deduction. If so, human individuals are not the wet-wear for deductive logic, at least in the versions that have surfaced in serious ways in the sprawling research programmes of modern logic. There is a particularly interesting reason for this. If we ask what the value of deductive consequence is, the answer is that it is a guarantee of truth-preservation. Guaranteed truth-preservation is a guaranteed way of avoiding error.[9] But individual agents are not in the general case dedicated to error-avoidance. So for the most part the routines of deductive consequence do not serve the individual agent in the ways in which he is disposed (and programed) to lead his cognitive and decisional life. This is not to say that agents do not perform deductive tasks even when performing on the ground level of our hierarchy. There is a huge psychological literature about such behaviour (accessibly summarized in Manktelow [1999]). The point

[8]In fact, it is better thought of as *minimalist* rationality, the rationality involved in making do with scarce resources.

[9]That is, of avoiding errors not alredy in his data-base or his premiss-set.

rather is that deductive thinking is so small a part of the individual's reasoning repertoire.

Indispensable to agency is the ability to remember. The literature on memory recognizes a contrast between *occurrent* and *dormant* memory, echoing a distinction between short term and long term memory.[10] These two reminisciential operations work in interestingly different ways. Occurrent memory presents beliefs that are, here and now, ready for action, for driving inferences and shaping behaviour. On the other hand, beliefs stored in dormant memory are not accessible as premisses in inferences; they do not conflict with or interact with one another; and some researchers are of the view that they do not influence conduct.

Occurrent memory is governed by tougher requirements than dormant memory. Occurrent or short term memory is in some sense bothered by inconsistency, whereas inconsistencies in dormant memory are virtually inert. This difference also crops up in the following way. Occurrent inconsistency is something a rational agent will, in one way or another, try to do something about. Dormant inconsistences tend not to register in an agent's consciousness. By and large there is nothing to *be* done about them.

Beliefs and memories are not the only things held to consistency assumptions in the lives of individual agents. Desires are also commonly expected to be consistent; "beliefs and desires can hardly be reasons for action unless they are consistent" (Elster [1985, 4]). In present-day social science, consistency is often defined for *preferences*. Transitivity is the minimal condition on preference. If agent S prefers X to Y and Y to Z then he can be expected also to prefer X to Z.

There is reason to doubt this assumption. "To recognize... inconsistency'[of beliefs or preferences] does not demolish [one's economic theory], since both the feasibility and the necessity of "consistency" also requires justification... Whatever view one takes of, say Agamemnon's dilemma, it can scarcely be solved by simply demanding that Agamemnon should lick his preference ordering in shape before he gets going" (Sen [1987, 66]). While it is quite true that people usually try to adjust their preferences if intransitivities are pointed out to them, the fact is that adopting intransitive preferences is far from uncommon. Moreover, even after intransitivities have been identified, some people persist in their preferences (Raiffa [1968, 75]).

[10]See Howe [1970], Collins and Quillam [1969] and Lindsay and Norman [1977]. For the classic studies see Howe [1970], Collins and Quillam [1969] and Lindsay and Norman [1977].

Time preference inconsistency is also a frequent, even endemic, occurrence in human life. People are strikingly prone to non-exponential time preferences, that is, to time preferences in which the future is discounted at non-constant ratios. However, here too it seems that remedies exist which enable the non-exponential time preferer to deal with this problem. George Ainslie has suggested that by collecting "several future choices the chances are increased that in each of them one will take the option with a later and greater reward" (Elster [1983] and Ainslie [1982]).

The principle that preference should be continuous is breached by what are called non-Archimedean preferences, such as lexicographic preferences attached to normative hierarchies. Here is an informal example, which we quote from Elster [1983, 9].

> If I am starving and am offered the choice between an option involving one loaf and listening to a Bach record and another involving one loaf and listening to Beethoven, then my love for Bach may make me prefer the first option. If, however, from the first option is subtracted even a very small crumb of bread, as small as you please, then I switch to the second because at starvation level calories are incomparably more important than music.

We see then, that the continuity principle is not just breached, it is refuted. We might wonder why it was proposed in the first place. The answer is that such an assumption helps streamline theory. If preference is held to the conditions of transitivity, completeness and continuity, preferences are representable as real-valued utility functions and the mathematics of preference manipulation takes on a certain power and elegance. The assumption guarantees that the economic theory will not be descriptively adequate. The question is whether it is a defensible idealization of an individual's behaviour; can the supposed regularity be considered to have normative force? (For a negative answer, see Woods [2001d, Chapter 8]. On the more positive side, see Johnson [2000].)

6 PRACTICAL LOGICS

In our description of it so far, we have left the theory of practical reasoning a fairly underdetermined affair. There is a desirable utility in such flexibility. We leave ourselves free to consider the pros and cons of extending or adapting our protologic in many possible ways, and in so doing availing ourselves of the benefit of work already done and on the record. There is a lot of it, too, whether temporal logics (e.g., van Benthem [1991]), logics

of action (e.g., Davidson [1980], Brand and Walton [1975], Brand [1984]), dynamic logic (e.g. van Benthem [1996]), not to forget the huge literature on deontic logic, and the practical logics of pragmatic philosophers (e.g., Dewey [1938] and Schiller [1929]).

There are multiples of different ways of finishing a theoretical product from its relatively modest beginnings as a logic supplemented by designated resources for the treatment of action and time. This leaves the research community with multiples of chances of coming up with finished products that receive and deserve consensus of a sort that we do not yet see much in evidence. Even so, it is an attraction of our protologic that it serves the desirable end, and achieves the welcome economy, of a principled and modest shortening of the list of attributes on whose behalf the adjective "practical" is invoked. If we return to the list developed in Section 3 of the present chapter, it is clear that our protologic sanctions some deletions.

A practical logic in our sense is not restricted to the study of reasoning about ordinary or commonplace matters. Nothing precludes the practical reasoner rushing to finish an arcane proof under press of his publisher's deadline.

A practical logic in our sense is no enemy of the alethic. For example, there is a well-understood role in dialogue logic for parties to enhance their shared data-bases. In so doing they increase their resources for making more direct cases for various actions.

Practical logic pertains to moral reasoning but is not restricted to it. Nor does it exclude factual reasoning. (See above).

Practical logic is no enemy of formality. Where appropriate it can involve express manipulation of logical forms; and even where reasoning is not formal in so sharply structural a way, practical logic is amenable to other grades of formal treatment. (Woods [1980], [1989], [2001e, Chapter 15], van Eemeren et al. [1996], cf. Johnson [1996, 120]).

Practical logic is not inherently about fuzzy reasoning, but can be extended to a fuzzy logic (e.g., Zadeh [1975], Chang and Lee [1975], Lee [1972], Przelecki [1976] and Hájek [1998]) or to a logic of vagueness (e.g. Tye [1990], Williamson [1994]) in those cases in which reasoning requires attending to in a more or less direct way the fuzziness of terms or, to fuzzy states of affairs. There are those who argue that practical reasoning is inherently fuzzy in just this sense. In our view this is an open question. (See, e.g., Woods [2000f].)

Practical logic subsumes but is not restricted to what Aristotle calls practical syllogisms. The same is true for the adaptation of the same idea in Gabbay and Woods [1999]. In a practical logic of the kind under review, a move in a dialogue always occasions an action by the other party, even though his action needn't be the action, if any, implied or suggested by his *vis-à-vis* premisses. For example, one party may say to the other: "So, you see, you ought to mow the lawn now". One way for the second party to

react to that move is to start mowing the lawn. This is an explicit action that will also serve as implicit acceptance of his interlocutor's claim. Or he might reply, "Yes, I really should be mowing the lawn," which is explicit acceptance and intimation of an action yet to be taken. A third answer is "Like hell!" which is an explicit (and emphatic) rejection. A fourth is phoning a friend to arrange for a golf game, which is explicitly not mowing the lawn and implicit rejection of the argument that called for it.

Practical logic deals with goal-directed, purposive action, but not at the expense of context-free approaches. One of the central tasks of a logic of propositions is the construction of a proof theory whose rules can be seen as virtual routines discharged by arbitrary agents.

Neither do we think that practical logic should be reserved for reasoning involving incommensurabilities. Incommensurability is ambiguous (Gray [2000]). In its most basic sense, reasoning from incommensurabilities is reasoning of a *pluralistic* kind. It is illustrated by the following schema.

1. Harry and Sarah value both friendship and patriotism.
2. Friendship and patriotism though different, and sometimes behaviourally non-co-satisfiable, are incomparable values.
3. In circumstances K, Harry opted for friendship and Sarah for patriotism.
4. Both acted rightly. Period.

It is true that normative reasoning is often occasion for judgements of incommensurability, but this is also sometimes times of scientific thinking. Pluralism abounds in logic, for example. And paraconsistent logics have been purpose-built to accommodate incommensurabilities (in the form of outright inconsistencies) whether in set theory or quantum mechanics. (Priest [1998], Brown [1993]) However, the incommensurability view of practicality intersects with our own conception, in the following way. Sometimes when faced with an incommensurability or an inconsistency, the practical (i.e., individual) agent has no realistic option but to let it be. He may lack the resources to adjust his data-base for consistency, which puts him in a situation in which he must think or act *in spite* of inconsistency. On the other hand, the very resources that an individual agent sometimes lacks are progressively available to agents of higher type.

The only interpretation that we ourselves are able to give the applied versus theoretical distinction in practical logic is one of the following inequivalent pair. First is the distinction between reasoning in a fully interpreted as opposed to a merely semi-interpreted vocabulary. To achieve its generality economically, a practical logic may operate with a semi-interpreted object language. But it will also have the means of giving its theorems full interpretations. (This is tricky. No such procedure will preserve formal

invalidity. See Woods, Chapter 2, below) The second way of drawing our present distinction is to see it as an instance of a particular way of construing the descriptive-normative distinction. In a widely accepted view of this latter, the task of finding a descriptive application of a normative theory is a matter of (a) finding the discrepancies between them, and (b) accounting for the descriptive deviations as approximations to the ideal conditions, full compliance with which would qualify as normatively perfect performance.

Unless we are mistaken, the sense we have proposed to give our proto-logic offers guidance on the applicability of other distinctions appropriated by those intent on giving "practical" some principled meaning. The purported distinction between concrete and abstract is handled by what we have said about the applied-theoretic distinction. Also there covered is the distinction between unregimented language and canonical notation. The distinction between a natural logic and an artificial logic can be captured by the distinction just mentioned. Alternately it is the distinction between the psychologically real and the psychologically ideal, which we have already discussed.

There is also an intuitive distinction between tasks whose performance requires little or no tutelage and those whose performance require specialized technical information. Cutting across this distinction, but in ways that produce some degree of overlap, is the contrast between ordinary and esoteric subject matters. If we wanted the distinction between practical and theoretical logics to be constrained by these contrasts, they would push in somewhat different directions; and formal logics such as *FDL* would elude classification altogether. We ourselves see little appeal in the first of these proposed criteria. A logic that attempted to give some insight into what goes on when an individual attempts to solve the Four Colour Problem is, as much a practical logic as any that attempted to elucidate an agent's choice of breakfast cereal. Neither are we persuaded that, for our purposes here, there is any abiding value in the contrast between the ordinary and everyday and (say) the business of quantum nonlocality in physics. A more fruitful way of drawing the contrast between a practical and theoretical logic is by piggy-backing on our distinction between a practical and a theoretical agent. The value of so doing (apart from the naturalness of the concurrence) is that it is very much less necessary to discredit a logic for its failure to model realistically actual human behaviour. Most mainstream logic since 1879, and most direct rivals of it, are subject to this failure. They fail for the most part because their strategies are too complex for the computational capacities of human individuals or, because their latitude in other respects (e.g., monotonicity), exceeds actual human reach. True, some mitigation of these misrepresentations can be found in the notion of idealization; but idealization is a more fraught device than is usually recognized (one cannot idealize at will). Even so, many of these logics, which fail as principled descriptions of what human individuals are capable of, succeed or come

closer to succeeding as formalized accounts of what institutional agents are capable of. So a decision to regulate the distinction between practical and theoretical logics in this way has the virtue, even on an idealized agent approach to logic, of saving much of what fails as a practical logic as what succeeds as a theoretical logic.

A final word about user-friendly logics. The idea can be given both a weak and a strong interpretation. On the weak interpretation, a user-friendly logic is, as Woods and Walton once said, "a manual of self-help for the ratiocinatively insecure" (Woods and Walton [1989, xvii]). So conceived, a logic is user-friendly to the extent that it can with little or no formal instruction be applied to the analysis and evaluation of real-life arguments. We would hope that good theories of practical reasoning would score at least moderately well on the count of user-friendliness. By this we mean that it is a virtue to avoid unnecessary technicalities and dispensable complexities. On the other hand, we have little affection for user-friendliness on its strong interpretation. In this meaning, a theory is user-friendly to the extent that its authoritative claims (its theorems, so to speak) are discernible without formal instruction to the untutored eye of the ordinarily competent reasoner. As we see these things, theories that have highly counterintuitive mathematical structures can turn out to be the right theories, and can in turn be attended by application procedures which make them highly user-friendly on the weak interpretation. (Indeed, the grammar of English is one such theory.)

We have already said that we find ourselves somewhat vexed by the descriptive-normative distinction in logic. As we bring this section to a close, it would be helpful if we could briefly shed some light on our reservation.

There is a considerable body of opinion in the century and a quarter since 1879 that a logician's job is axiomatization and that axioms are what the logician finds to be most intuitive. Much the same view can be found among logicians who favour natural deduction approaches. Here, too, one's choice of structural and operational rules is seen as a matter of what strikes the theorist as most intuitively correct. Much the same *modus operandi* is evident in other disciplines, especially abstract disciplines that lack in any direct way anyhow — empirical checkpoints. In philosophy this approach is the heart and soul of conceptual analysis in the manner of G. E. Moore and an entire generation which fell under his influence.

The method of analytic intuitions raises a fundamental methodological question. Given that an intuition is what the theorist antecedently believes, and that a fundamental intuition is what he believes utterly, is there any good reason to suppose that intuitions are *epistemically privileged*? Is there any reason to suppose that what the theorist believes utterly qualifies as knowledge? If the answer is Yes, the essential methods of conceptual analysis are confirmed. If the answer is No, the methodology of the abstract sciences must take this into account.

One attraction of the method of analytic intuitions in logic is that it secures a comfortable purchase on the shelf of normativity. It allows for it to be the case that a human being *should* reason in such-and-such a way, if the logician-theorist's intuitions lend support to a rule or a theorem to the same effect. But shorne of the comforts of the method of analytic intuitions, the normatively minded logician will find less desired normativity a lot more difficult to get a sure grip on. It may be that such a theorist would be well-served in taking the following approach.

> First, he might try to make this account conform closely to how in the general case practical agents actually perform under the conditions the theory takes note of.
>
> Secondly, he might also try to take note of what in actual practice is regarded as mistakes or errors.

If he does both these things, we will say that his account is *descriptively adequate*. The sixty-four dollar question is this:

> Can the theorist obtain a serviceable standard of normativity by putting it that a practical agent performs as he should if his performance conforms to what his fellows do and is not marred by mistakes in the sense of a paragraph ago?

This we leave as an open question. (But see Woods [2001d, Chapter 8].)

There is an ancient way of characterizing the practical. It is to be found in the contrast between Practical and Theoretical Reason, between *phronesis and epistēmē*. Perhaps we now have the wherewithal to characterize this contrast in ways that would be found credible by present-day readers. Accordingly, we repeat our proposal that Practical Reason be thought of as a repertoire of skills characteristic of the lower strata in the hierarchy of agency, that Theoretical Reason be thought of as sets of skills characteristic of higher up, and that the contrast be seen as a matter of degree — a matter of how low down and how high up the agent in question chances to be. Here is a suggestion which preserves the truth that *all* reasoning is goal-directed, that all reasoning portends *some* kind of action. But it allows us to crosscut this universality with considerations of indigenous import, in which Practical Reason is characterized by features of the *agent* whose reasoning it is.

It is also well to emphasize that we are taking the agency view of logic, as opposed to the linguistic view. The distinctions we have been tracking and the exclusions we have been proposing, have been transacted within the tent of agency logic. Agency logic is the natural home of practical logic, and offers reasonable accommodation to one reasonable conception of theoretical logic. It is not our view that the linguistic conception of logic should be rejected. There is nothing good to be said for the idea that we should say

no to recursion theory, model theory, proof theory and set theory. This is a *Handbook* about the practical turn in logic. It obliges us to give sense to what is practical and to give some idea as to where the idea of the practical is best pursued by logical theory. In the end, it is this question which we bring to the distinction between the agency and linguistic conceptions of logic. And, with respect to the matters that concern us here, it is our view that an agency logic is a natural home for practical reasoning and that a linguistic logic is not. But saying so is a long way from pleading the exclusion of linguistic logic.

7 ALLIED DISCIPLINES

In absorbing the dialogical approach to practical reasoning, we are free to engage — to appropriate or adapt — a large research literature. Dialogue logics come in a variety of stripes, some of the most interesting of which are Hamblin [1970], Lorenzen and Lorenz [1978], Barth and Krabbe [1982], Carlsen [1982], MacKenzie [1990], Walton and Krabbe [1995], Girle [1993], [1996], [1997], and Gabbay and Woods [2000] and [2000b]. A bounty of rich resources also extends to developments in cognitive science, AI and linguistics.

We take it as obvious that, irrespective of how we finally settle the question of the normative-descriptive distinction for theories of practical reasoning, it would be a mistake to ignore developments in these allied disciplines. For example, consider the impact of psychology. The psychological studies to date have concentrated on deductive, and probabilistic and inductive reasoning, with somewhat less attention given to decisional and causal reasoning. There is no simple dominant paradigm at present; in fact, there are at least four main approaches that are currently in contention. These are the *mental models* account (e.g., and Byrne [1991]), *mental logics* (e.g., Rips [1994]), *rational analysis and information gain* (e.g., Chater and Oaksford [1999], Oakford, Chater, Grainger and Larkin [1996]), and *domain specific reasoning schemas* (e.g., Evans and Over [1996]). Notwithstanding these theoretical and methodological differences, experimental evidence bears on the business of practical reasoning in two especially telling ways. One is that human beings do indeed seem disposed to commit fallacies, that is, errors of reasoning which are widely and cross-culturally made, easy to make and attractive, and difficult to correct. (Woods [1992]). A second point is that human reasoning performance seems to improve, that is, to commit fewer fallacies, when the reasoning in question is set in a deontic-context. (Cheng and Holyoak [1985]). "Deontic" here means directed to or (productive) of an action, which is the core sense of our notion of practicality. Since our protologic is already moored in deontic and prudential contexts, a mature theory which is an extension of it must try to explain what is and what

isn't a fallacy in a deontic environment or in a practical reasoning task, and why impractical reasoning should be more prone to fallacies than practical reasoning. It is entirely possible that some of this difference lies in the fact that one and the same strategy might be a reasoning error in a nonpractical context of reasoning, and yet be an error-free strategy deontically. (Gabbay and Woods [1999], Gabbay and Woods Chapter 11, below.)

A practical logic should also incorporate important developments in the AI sector. It should exploit the fact that human reasoning is non-monotonic and that non-monotonic structures have been investigated by AI researchers (e.g., Geffner [1992] and Pereira, chapter 10, this volume). Human reasoners are also adept at recognizing and manipulating defaults. A default is something taken as true provisionally or, as is said, in default of information to the contrary. (Reiter [1980]) Default reasoning introduces into the business of human inference some extraordinary economies, which a practical logic must take pains with. For reasoning is good not only when it produces the right answer, but when it produces it on time. As a related development from linguistics, generic inference discloses its thinking to default reasoning. Generic claims are generalizations of a particularly robust kind. Like "Tigers are four-legged," they tolerate true negative instances. (Carlson and Pelletier [1995]) They also seem triggered by very small samples, as we have seen. The two features are linked. Somehow human beings are rigged for what classically would be seen as hasty generalization fallacies in precisely these cases in which the reasoner is not generalizing to a universally quantified conditional (which is as fragile as a generic generalization is robust), but rather to a generalization certain negative instances of which happen not to matter.

It is easy to see how default reasoning and generic inference touch on the classical fallacy of hasty generalization, and necessitate a substantial reconsideration of its traditional analysis. Other forms of default reasoning pertain in the same way to the classical fallacy *argumentum ad ignorantiam*. The basic structure of the fallacy is the argument form:

1. It is not known that P
2. Therefore not P.

On the standard analysis, *ad ignoratiam* arguments are not only deductively invalid, but wholly implausible as well. But as studies of autoepistemic reasoning show (e.g.,) there are exceptions to so harsh a verdict, as witness:

1. If there were a Department meeting today, I would know about it.
2. But in fact I know nothing of any such meeting.
3. So, there'll be no meeting.

Here is further occasion for a mature theory of practical reasoning to winnow out the mistakes in classical accounts of fallacious reasoning. (Concerning which see Gabbay and Woods [2003])

8 SUMMARY

1. By now we share the widely received view that *FDL*, or standard deductive logic is not an adequate logic of practical reasoning or argument.

2. However, we see little virtue in iconoclasm. *FDL* has secured an impressive place in our intellectual heritage. It is perhaps most fundamentally seen as a theory of the consequence relation. So conceived of, it is wholly intelligible why *FDL* would not have much success as a theory of inference or argument. No one doubts that there is more to reasoning and arguing than manipulating the consequence relation. On the other hand, what theory of reasoning or argument can be *indifferent* to the consequence relation. So we propose an ecumenical attitude towards *FDL*.

3. More particularly, we recognize both the agency and the linguistic conceptions of logic. Accordingly, we propose a flexible use for the word "logic". A logic is a theory or reasoning or a theory of argument which makes a reasonable claim on satisfying one of this *Handbook*'s organizing themes — "Systematic and Foundational Aspects".

4. The emphasis of the *Handbook* is practical reasoning. Accordingly, a practical logic is any suitable extension of our protologic, which is a logic of action in real time, whose further properties are indicated by the type of agency involved. We could say that a purely practical protologic is one that recognizes the resource limitation strategies at the base of our hierarchy of agents.

5. We ourselves are not much drawn to embedding the protologic in a dialogue logic or family of dialogue logics. In our view it should be the other way around. Dialogical reasoning is a paradigm of social reasoning which for us is important but not exhaustive. Dialogical reasoning is sometimes easier to analyse than purely private, solo reasoning. Even so, a good many of the principles of dialogue logic can be expected to adapt in a fairly direct way to solo contexts. So we propose dialogue logic as a species of practical logic.

6. A practical logic should take note of, and try to account for, the disposition of humans to commit fallacies, and the diminished disposition to commit fallacies in deontic contexts.

7. Owing to the difficulties involved in regulating the normative-descriptive distinction in a principled and illuminating way, the present authors propose descriptive adequacy as a prime target of a practical logic. (However, we are aware that not all of our authors find the distinction as perplexing as we do.)

BIBLIOGRAPHY

[Ainslie, 1982] G. Ainslie. A behavioral economic approach to the defense mechanism: Freud's energy theory revisited. *Social Science Information*, 1982.

[Arnauld and Nicole, 1996] Antoine Arnauld and Pierre Nicole. *Logic or the Art of Thinking*. Cambridge: Cambridge University Press, 1996. Originally published in 1662. This volume edited by Jill Vance Buroker.

[Axelrod, 1984] Robert Axelrod. *The Evaluation of Cooperation*. New York: Basic Books, 1984.

[Axsom *et al.*, 1987] D.S. Axsom, S. Yates, and S. Chaiken. Audience response as a heuristic case in persuasion. *Journal of Personality and Social Psychology*, 53:30–40, 1987.

[Bacon, 1960] Francis Bacon. *The New Organon and Related Writings*. New York: The Liberal Arts Press, 1960. Fulton H. Anderson (ed.).

[Barth and Krabbe, 1982] E.M. Barth and E.C.W. Krabbe. *From Axion to Dialogue: A Philosophical Study of Logic and Argumentation*. Berlin: de Gruyter, 1982.

[Brand and Walton, 1975] Myles Brand and Doulgas Walton. *Action Theory*. Dordrecht: Reidel, 1975.

[Brand, 1984] Myles Brand. *Intending and Action: Toward a Naturalized Action Theory*. Cambridge MA: MIT Press, 1984.

[Bratman, 1999] Michael Bratman. *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge: Cambridge University Press, 1999.

[Brown, 1993] Bryson Brown. Old quantum theory: A paraconsistent approach. *Proceedings of the Biennial Meetings of the Philosophy of Science Association*, 2:397–411, 1993.

[Carlsen, 1982] Laurie Carlsen. *Dialogue Games*. Dordrecht and Boston: Reidel, 1982.

[Carlson and Pelletier, 1995] Gregory N. Carlson and Francis Jeffry Pelletier. *The Generic Book*. Chicago: Chicago University Press, 1995.

[Change and Lee, 1975] C.L. Change and R.C.T. Lee. Some properties of fuzzy logic. *Information and Control*, 19:417–431, 1975.

[Chater and Oaksford, 1999] N. Chater and M. Oaksford. The probablity heuristics model of syllogistic reasoning. *Cognitive Psychology*, 39:191–258, 1999.

[Cheng and Holyoak, 1985] P.W. Cheng and K.J Holyoak. Pragmatic reasoning schemas. *Cognitive Psychology*, 17:391–416, 1985.

[Cherniak, 1986] Christopher Cherniak. *Minimal Rationality*. Cambridge, MA: MIT Press, 1986.

[Church, 1956] Alonzo Church. *Introduction to Mathematical Logic*. Princeton, NJ: Princeton University Press, 1956.

[Churchland, 1981] Patricia Churchland. On the alleged backwards referral of experiences and its relevance to the mind-body problem. *Philosophy of Science*, 48:165–181, 1981.

[Coady, 1992] C.A.J. Coady. *Testimony: A Philosophical Study*. Oxford: Oxford University Press, 1992.

[Collins and Quilliam, 1969] A. Collins and M. Quilliam. Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8:240–247, 1969.

[Davidson, 1980] Donald Davidson. *Essays on Actions and Events*. Oxford: Oxford University Press, 1980.

[Davidson, 1984] Donald Davidson. *Inquiries into Truth and Interpretation*. New York: Oxford University Press, 1984.

[Dennett, 1996] Daniel C. Dennett. *Kinds of Minds: Toward an Understanding of Consciousness.* New York: Basic Books, 1996.

[Dewey, 1938] John Dewey. *The Theory of Enquiry.* Vol 12 of JoAnn Boydston (ed.), *John Dewey: The Later Works, 1925–1953*, Carbondale, IL: Southern Illinois University Press, 1981, 1938.

[Dummett, 1973] Michael Dummett. *Frege: Philosophy of Language.* London: Duckworth, 1973.

[Dunn, 1984] J. M. Dunn. Relevant logic and entailment. In D. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume III, pages 117–224. Dordrecht: Reidel, 1984.

[Eagly and Chaiken, 1993] A. H. Eagly and S. Chaiken. *The Psychology of Attitudes.* Fort Worth: Harcourt Brace Jovanovich, 1993.

[Elster, 1983] Jon Elster. *Sour Grapes: Studies in the Subversion of Rationality.* Cambridge: Cambridge University Press, 1983.

[Elster, 1985] John Elster. *Sour Grapes: Studies in the subversion of rationality.* Cambridge: Cambridge University Press, 1985.

[Evans and Over, 1996] J.St. B.T. Evans and D.E. Over. *Rationality and Reasoning.* Hove, UK: Psychology Press, 1996.

[Frankfurt, 1988] Harry Frankfurt. *The Importance of What We Care About.* Cambridge: Cambridge University Press, 1988.

[Gabbay and Woods, 1999] Dov M. Gabbay and John Woods. Cooperate with your logic ancestors. *Journal of Logic, Language and Information*, 8, 1999.

[Gabbay and Woods, 2000] Dov M. Gabbay and John Woods. Non-cooperation in dialogue logic: Getting beyond the goody two-shoes model of argument. *Synthese*, 2000.

[Gabbay and Woods, 2000a] Dov M. Gabbay and John Woods. Editorial. *Journal of Logic and Computation*, 10(1):1–2, 2000a.

[Gabbay and Woods, 2000b] Dov M. Gabbay and John Woods. More on non-cooperation in dialogue logic. *Logic Journal of the IGPL*, 2000b.

[Gabbay and Woods, 2000c] Dov M. Gabbay and John Woods. The new logic. *Logic Journal of the IGPL*, 2000c.

[Gabbay and Woods, 2002] Dov M. Gabbay and John Woods. *The Reach of Abduction.* 2002. To appear.

[Gabbay and Woods, 2003] Dov M. Gabbay and John Woods. *A Resource-Based Theory of Fallacies: A Study in Economic Rationality.* 2003. To appear.

[Geffner, 1992] Hector Geffner. *Default Reasoning: Causal and Conditional Theories.* Cambridge, MA: MIT Press, 1992.

[Girle, 1993] Roderic A. Girle. Dialogue and entrenchment. In *Proceedings of the Sixth Florida Aritifical Intelligence Research Symposium*, pages 105–189. Fort Lauderdale, 1993.

[Girle, 1996] Roderic A. Girle. Knowledge: Organized and disorganized. In *Proceedings of the International Conference on Formal and Applied Practical Reasoning*, pages 246–260. Bonn, 1996.

[Girle, 1997] Roderic A. Girle. Belief sets and commitment stores. In Hans V. Hansen, Christopher W. Tindale, and Athena V. Colman, editors, *Argumentation and Rhetoric.* St. Catharines, ON: Ontario Society of the Study of Argument, 1997.

[Gray, 2000] John Gray. *Two Faces of Liberalism.* New York: The New Press, 2000.

[Hájek, 1998] Petr Hájek. *Metamathematics of Fuzzy Logic.* Dordrecht and Boston: Kluwer, 1998.

[Hamblin, 1970] C.L. Hamblin. *Fallacies.* London: Methuen, 1970.

[Hintikka and Bachman, 1991] Jaakka Hintikka and James Bachman. *What if ...?* Mountain View, CA: Mayfield, 1991.

[Hintikka, 1989] Jaakko Hintikka. The role of logic in argumentation. *The Monist*, 72, 1989.

[Howe, 1970] M. Howe. *Introduction to Human Memory.* New York: Harper and Row, 1970.

[Jackson, 1996] Sally Jackson. Fallacies and heuristics. In *Logic and Argumentation*, pages 101–114. Amsterdam: North-Holland, 1996.

[Jacobs and Jackson, 1983] Scott Jacobs and Sally Jackson. Speech act structure in conversation: Rational aspects of pragmatic coherence. In Rober T. Craig and Karen Tracy, editors, *Conversational Coherence: Form, Structure, and Strategy*, pages 47–66. Newbury Park, CA: Sage, 1983.

[Jacobs *et al.*, 1985] Scott Jacobs, M. Allen, Sally Jackson, and D. Petrel. Can ordinary arguers recognize a valid conclusion if it walks up and bites them in the butt? In J.R. Cox, M.O. Sillars, and G.B. Walker, editors, *Argument and Social Practice: Proceedings of the Fourth SCA/FA Conference on Argumentation*, pages 665–674. Annandale, VA: Speech Communication Association, 1985.

[Johnson-Laird and Byrne, 1991] P.N. Johnson-Laird and Ruth M.J. Byrne. *Deduction: Essays in Cognitive Psychology.* Hove and London: Lawrence Erlbaum Associates, 1991.

[Johnson, 1996] Ralph H. Johnson. *The Rise of Informal Logic.* Newport News, VA: Vale Press, 1996.

[Johnson, 2000] Ralph H. Johnson. *Manifest Rationality: A Pragmatic Theory of Argument.* London: Lawrence Erlbaum Associates, 2000.

[Lee, 1972] R.C.T. Lee. Fuzzy logic and the resolution principle. *Journal of the Association of Computing Machinery*, 19:109–119, 1972.

[Lindsay and Norman, 1977] P. Lindsay and D. Norman. *Human Information Processing.* New York: Academic Press, 1977.

[Locke, 1975] John Locke. *An Essay Concerning Human Understanding.* Oxford: Clarendon Press, 1975. Originally published in 1689.

[Lorenzen and Lorenz, 1978] Paul Lorenzen and Kuno Lorenz. *Dialogische Logik.* Darmstadt: Wissenschaft-liche Buchgesellschaft, 1978.

[Mackenzie, 1990] J. Mackenzie. For dialogue systems. *Studia Logica*, XLIX:567–583, 1990.

[Manktelow, 1999] Ken Manktelow. Hove, UK: Psychology Press, 1999.

[Mill, 1974] John Stuart Mill. A system of logic. In J.M. Robson and J. Stillinger, editors, *The Collected Works of John Stuart Mill*, volume VII and VIII. Toronto, ON: University of Toronto Press, 1974. Originally published in 1843, London: Longman and Green. Volume VII was published in 1973 and Volume VIII was published in 1974.

[Minsky, 1975] Marvin Minsky. Frame-system theory. In R.C. Schank and B.L. Nash-Webber, editors, *Theoretical Issues in Natural Language Processing.* 1975. preprints of a conference at MIT, June 1975. Reprinted in P.N. Johnson-Laird and P.C. Wason (eds.), *Thinking: Readings in Cognitive Science*, pages 355–376. Cambridge: Cambridge University Press 1977.

[Oaksford *et al.*, 1996] M.R. Oaksford, N. Chater, B. Grainger and J. Larkin. Optimal data selection in the reduced array selection task (RAST). *Journal of Experimantal Psychology: Learning, Memory and Cognition*, 23, 441–458, 1997.

[O'Keefe, 1990] D.J. O'Keefe. *Persuasion: Theory and Research.* Thousand Oaks, CA: Sage, 1990.

[Peirce, 1992] C.S. Peirce. *Reasoning and the Logic of Things: The 1898 Cambridge Conferences Lectures by Charles Sanders Peirce.* Cambridge, MA: Harvard University Press, 1992. Kenneth Laine Ketner (ed.).

[Petty and Cacioppo, 1986] R.E. Petty and J.T. Cacioppo. *Communication and Persuasion.* New York: Springer-Verlag, 1986.

[Petty *et al.*, 1981] R.E. Petty, J.T. Cacioppo, and R. Goldman. Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, 41:847–855, 1981.

[Priest, 1998] Graham Priest. To be *and* not to be — that is the answer: On Aristotle on the law of non-contradiction. *Philosophiegeschichte und Logische Analyse*, 1:91–130, 1998.

[Przełecki, 1976] M. Przełecki. Fuzziness as multiplicity. *Erkenntnis*, 10:371–380, 1976.

[Raiffa, 1968] H. Raiffa. *Decision Analysis.* Reading, MA: Addison-Wesley, 1968.

[Reiter, 1980] Raymond Reiter. A logic for default reasoning. *Artificial Intelligence*, 12:81–132, 1980.

[Rips, 1994] Lance J. Rips. *The Psychology of Proof: Deductive Reasoning in Human Thinking.* Cambridge, MA: MIT Press, 1994.

[Rosch, 1978] Eleanor Rosch. Principles of categorization. In Eleanor Rosch and B.B. Lloyd, editors, *Cognition and Categorization*, pages 27–48. Hillsdale, N.J.: Erlbaum, 1978.

[Schiller, 1929] F.C.S. Schiller. *Logic for Use: An Introduction to the Voluntarist Theory of Knowledge*. London: Bell, 1929.

[Sen, 1987] Amartya Sen. *On Ethics and Economics*. Oxford: Basil Blackwell, 1987.

[Sidgwick, 1910] Alfred Sidgwick. *The Application of Logic*. London: Macmillan, 1910.

[Smith and Medin, 1981] Edward E. Smith and Douglas L. Medin. *Categories and Concepts*. Cambridge, MA: Harvard University Press, 1981.

[Tye, 1990] Micheal Tye. Vague objects. *Mind*, 99:535–557, 1990.

[van Benthem, 1991] Johan van Benthem. *Language in Action: Categories, Lambdas and Dynamic Logic*. Amsterdam: North-Holland, 1991.

[van Benthem, 1996] Johan van Benthem. *Exploring Logical Dynamics*. Stanford: CSLI Publications, 1996.

[van Eemeren *et al.*, 1996] Frans H. van Eemeren, Rob Grootendorst, Francisca Snoeck Henkemans, J. Anthony Blair, Ralph H. Johnson, Erik C.W. Krabbe, Christian Plantin, Douglas N. Walton, Charles A. Willard, John Woods, and David Zarefsky. *Fundamentals of Argumentation Theory: A handbook of historical backgrounds and contemporary developments*. Mahwah, NJ: Erlbaum, 1996.

[Velleman, 2000] J. David Velleman. *The Possiblity of Practical Reason*. Oxford: Clarendon Press, 2000.

[Walton and Krabbe, 1995] Douglas Walton and E.C.W. Krabbe. *Commitment in Dialogue*. Albany, NY: SUNY Press, 1995.

[Whately, 1848] Richard Whately. *Elements of Logic, 1826; 9th edition*. London: Longmans, 1848.

[Williamson, 1994] Timothy Williamson. *Vagueness*. New York: Routledge, 1994.

[Wolfram, 1984] Stephen Wolfram. Computer softwear in science and mathematics. *Scientific American*, 251:188, September 1984.

[Woods and Walton, 1989] John Woods and Douglas Walton. *Fallacies: Selected Papers 1972-1982*. Berlin and New York: Foris de Gruyter, 1989.

[Woods, 1980] John Woods. What is informal logic? In *Informal Logic: The First International Conference*, pages 57–68. San Francisco: Edgepress Publishers, 1980.

[Woods, 1989] John Woods. The necessity of formalism in informal logic. *Argumentation*, 3:149–167, 1989. Reprinted in *Death of Argument: Fallacies and Other Seductions*published by Vale Press to appear in 2001. John Woods editor.

[Woods, 1992] John Woods. Public policy and standoffs of force five. In E.M. Barth and E.C.W. Krabbe, editors, *Logic and Political Culture*, pages 97–108. Amsterdam: North-Holland, 1992.

[Woods, 2000f] John Woods. Has informal logic anything to learn from fuzzy logic? In Hans V. Hansen and Christopher Tindale, editors, *Argumentation at Century's End*. St. Catharines, ON: OSSA, 2000f. To appear.

[Woods, 2001d] John Woods. *Paradox and Paraconsistency: Conflict Resolution in the Abstract Sciences*. Cambridge: Cambridge University Press, 2001d. To appear.

[Woods, 2001e] John Woods. *Death of Argument: Fallacies and Other Seductions*. Newport Beach: Vale Press, 2001e. In press.

[Woods, 2001f] John Woods. *Aristotle's Earlier Logic*. Paris: Hermes Science Publications, 2001f.

[Woodworth and Sells, 1935] R.S. Woodworth and S.B. Sells. An atmosphere effect in formal syllogistic-reasoning. *Journal of Experimental Psychology*, 18:451–460, 1935.

[Zadeh, 1975] L.A. Zadeh. Fuzzy logic and approximate reasoning. *Synthese*, 30:407–428, 1975.

[Zimmermann, 1989] Manfred Zimmermann. The nervous system and the context of information theory. In R.F. Schmidt and G. Thews, editors, *Human Physiology*. Berlin: Springer-Verlag, 2nd edition, 1989.

JOHN WOODS

STANDARD LOGICS AS THEORIES OF ARGUMENT AND INFERENCE: DEDUCTION

1 INTRODUCTION

This *Handbook* is aimed at a substantial audience of researchers and students whose work collectively represents several disciplines. Among them are logic, computer science, cognitive and experimental psychology, argumentation theory, forensic science and linguistics. Also involved are those branches of mathematics which explore Bayesian statistics, information theory and ergodics; engineering in its connection with complexity theory and entropy; economics and management studies in their engagement of game theory, choice theory and probabilistic decision theory; epistemology in its involvement with the foundations of probability and rationality theory; and fallacy theory and other branches of informal logic, as they bear upon the question of limitations (real or imagined) of formal logic as central and load-bearing part of any good theory of argument and inference, especially when considered as branches of *practical reasoning*. The aim of the present chapter and the one to follow is to give the untutored reader, or the reader who has long since forgotten his one course in the subject, an overview of standard logic, and to do so with sufficient rigour and in sufficient detail to qualify as a kind of crash course in logic at a level appropriate to the sophistication of the reader's own research programme in the theory of argument and inference. Of course, if the reader is a working logician, these pages can be breezed through quickly out of interest, or ignored altogether out of an impatience for the less standard chapters to come. Even so, I have tried to make these chapters worth pausing over even by the veteran logician. They include some essential background on Aristotle as a deductive logician; and they attend to the psychological relevance of deductive and inductive logic to recent discoveries in complexity theory.

For the purposes of these two chapters the term "logics" adumbrates a loose distinction between *deductive* and *inductive* systems. "Standard" suggests a motley of meanings, from "widely-received", "dominant", "best", and even "uniquely correct". It is best to give some backbone to this imprecision by stipulating that the systems here under review are those that have either acquired or approximated to the status of twentieth century orthodoxies. Even so, the concept of a standard logic is a slack and contentious idea. Its slackness is evidenced by the fact that even a more or less constant logical theory can be produced under significant variation of ax-

ioms, conditions, rules and means of expression.[1] And the contentiousness of the general concept is evidenced by the opinion of those theorists who hold, for example, that a logic of induction is not possible, (Popper [1992]: cf. Putnam [1975d], [1975b]).

If the basic idea of a standard logic lies open to imprecision and disagreement, the same is true of the suggestion that logic is a theory of argument and inference. This is perhaps best illustrated by the present state of formal *deductive* logic (*FDL*) which, for the purposes at hand, can tentatively be identified with the predicate calculus or, more narrowly, with the predicate calculus of first order (see below).

If, as a good many of its practitioners aver, the fundamental objective of standard deductive logic is to abet the recognition of a particular class of sentences, namely, the *logical truths*, (Quine [1950, xi]); or is, as others hold, to capture the properties of a propositional relation such as *entailment* or *logical consequence* (Dummett [1973] and Koslow [1992]), then it is not obvious that such logics will serve as good accounts of either argument or inference. Granted that there is more to argument and inference than any deductive logic could capture, there is room, even so, to resist the idea that a deductive logic will do equally well as a theory of *deductive argument* and *deductive inference* alike.

There are in such reservations two discernibly different points. One is that in as much as deductive logic is preoccupied with either the specification of logical truths or the characterization of logical implication or entailment, it must fail as an account of argument and of inference to the extent that neither the concept of a logical truth nor the concept of entailment will bear anything like the full weight of such multi-faceted structures as argument or inference. We might think of the point this way. Suppose that the would-be theorist of argument were told that his task is to give an account of good or correct arguments but that he must do this with an inventory of conceptual tools restricted to the concept of logical truth and others definable from it. According to our present critic, the theorist would have no prospect of capturing the idea of good or correct argument beyond specifying a notion of deductive correctness or validity. Fine as far as it goes, it would still fail, with an exception to be noted, to pick out a class of good arguments, even of good arguments aspiring to the standard of deductive correctness. The exception is the class of arguments meeting the condition of deductive validity whose premisses are true. Such arguments are called sound. Yet it may be doubted whether soundness is a sufficient condition on good arguments, even of those that aspire to be deductively good, as witness the *circular* but sound argument,

[1]For example, the truth of the sequent $\Phi_1, \ldots, \Phi_n \;/\; \Psi$ is independent of the language in which $\Phi_1, \ldots \Phi_n$ and Ψ are expressed. (Hodges [1983, 17]). Not all systems have this property, as witness Dunn and Belnap [1968].

Either A or not-A
Therefore, either A or not-A

and the *trivial* but sound argument (where "A" and "B" are true)

Both A and B
Therefore, B

Whether upon reflection these objections might be thought a decisive, or even a modest discouragement of the view that deductive logic has resources enough to be a good core theory of a certain conception of correct argument, deductive logic's prospects as a theory of inference run into heavier weather. Not all inferences are intended to be held to standards of deductive impeccability, but some are. Restricting our attention to those that are, it is unlikely that a theory of logical truth or of logical implication comes close to fulfilling the main criteria of inference-adequacy. It depends, of course, on the sort of thing the theorist takes inference to be pre-theoretically (just as the criticisms above need to be assessed in the context of how "argument" is understood pre-theoretically). But if we suppose inference to be a kind of belief-revision or updating of a cognitive agent's store of commitments, it is a matter of doubt as to whether the rules of deductive logic serve adequately as a constraint on such processes (Harman [1986], and his chapter 4 below; cf. Hintikka *et al.*, chapter 7 below). The deductive rule *modus ponens* is a case in point. If at a certain time t a reasoner holds that A, and if at $t + 1$ he comes to hold that if A then B, it is not invariably correct that at $t+1$ he should revise his commitments by adding B. He might at $t + 1$ come to see that B is false. His store of commitments is now threatened with inconsistency, and the reasoner might certainly avert it in more than one way, one of which might be correct. For example, he might now delete A or delete ⌜If A then B⌝. Either way, he is refusing the rule *modus ponens*, and the critic will say that, sometimes at least, doing so is precisely the right inference for him to make.

The proponent of *modus ponens* as an invariably correct rule of inference will be quick (and right) to notice that our example turns on a diachronic factor. There is a passage of time from t to $t+1$, and beyond. He may wish to insist that *modus ponens* is a correct rule of inference only under monochronic assumptions. That is, if at some same time a reasoner holds A and holds ⌜If A then B⌝, then it is always correct at that time to infer B, and if diachronicity be admitted at all, to infer B at any time just after any time at which he holds A and ⌜If A then B⌝ concurrently. The deductive theorist's reply comes to little more than this: If a reasoner holds at some t both A and ⌜If A then B⌝ and if there is nothing that contra-indicates the detachment of B just then, he should detach B. This is quite right, as trivialities have a way of being, but it is not *modus ponens*.

We shall return to this point in due course. Our purpose at present is to make the preliminary point that, for various reasons, the very idea of standard logics of argument and inference is a challenged notion, whatever its considerable appeal and its impressive historical bona fides. In the sections to follow, we shall present representative versions of such theories. In doing so, we shall make it a point to revisit these reservations from time to time.

Although our attention so far has been directed to systems of deductive logic considered as theories of argument and inference, reservations, some similar and some rather different, could be pressed against inductive logic. We do not here develop these objections, but reserve them for the subsequent chapter. It is enough for now to know that they exist and that they must eventually be taken into account. But in the little that we have said to date, it becomes apparent that *FDL* must be understood from the outset as taking some liberties with their respective subject matters. Two of these should be mentioned now. For one thing, contrary to what is often carelessly said, it is a rare thing to find a system of logic that purports to give anything but a *partial* account of its target concepts, though at times it may be offered as a complete account of that part of them. For another, it is entirely common for a theory of logic to misrepresent its subject matter, that is, to characterize it selectively and even to misstate certain of its features. In this, logic joins the company of virtually every other theory worthy of the name, whether mechanics, or population biology, or any other. It is sometimes regretted by critics that logic doesn't capture all the finely-grained texture of reasoning in real life, or of reasoning "on the hoof" as we might say. But it is naive to regret this as such. As with any theory worthy of the name, the regret is justified or not depending on what is left out and in what ways, and for what theoretical purposes real-life features are over- or underdescribed. These, too, are matters that we shall keep an eye on as we proceed, but it is well here to register a significant demurrer from the co-founder of the modern logic of quantification, who holds that there cannot be a logic of reasoning on the hoof, (Peirce [1992]).

2 HISTORICAL BEGINNINGS

It may seem odd, in a chapter devoted to 20th century orthodoxy in logic, that we should pause to say something about Aristotle. This we do not out of any sense of sentimentalism or misplaced historical enthusiasm. Aristotle is a neglected logician these days. More's the pity, since Aristotle's logical theory, especially the early account in the *Topics* and *On Sophistical Refutations*, manages to touch upon — often inadvertently — a good many of the issues dealt with in the present volume.

Aristotle originates his logical theory in the context of a richly developed tradition of argumentative practice nourished by two main arteries. On the

one hand there is the impressive record of Greek mathematics central to which were workable conceptions of *proof* and of *reductio* arguments. On the other, are the rhetorical traditions of the Sophists and the dialectical transformations of them developed in the writings of Plato. The word "dialectical" and its cognate "dialectic" is multiply-ambiguous in early Greek thought, as it is today in somewhat different ways. In the Socratic writings, the idea is no better illustrated than by the notion of refutation or *elenchus*. Refutations are a kind of *reductio* argument. The conceptual core of refutation is inducing an opponent to make concessions inconsistent with one of his stated claims or theses. For this to happen, it is not necessary, and in fact not desirable, that he be induced to say anything absurd or impossible, (and so they need not be *reductio ad absurdum* arguments). It suffices that what he says *now* cannot be true if what he said *then* (his original thesis) is true. Conversely, if what he says now is true, his original thesis is not true. It is easily seen that this form of argument closely resembles what Locke, over two millennia later, would call the *argumentum ad hominem*. *Ad hominem* arguments are those, said Locke, in which we "press a Man with Consequences drawn from his own Principles or Concessions" (Locke [1975, 686]).

Aristotle inherited this conception of argument and, with it, the problem of determining the conditions under which a refutation succeeded or failed. It is well to understand that Aristotle's problem arises in the context of everyday disputation, of arguments transacted by ordinary people about the great tangle of issues that people are exercised by. Also of importance is the point that not only are these arguments on the hoof, i.e., just as they come in natural speech uttered in real life, but they are *dialectical* and *informal* in senses that would be recognized and championed by many a present-day theorist of argument and informal logician. They are dialectical in the sense that they are interactive sequences of speech acts between two (or more) parties produced in real time;[2] and they are informal in the sense that for the purposes of their representation and assessment there is no pretense of restricting theoretical effort to the manipulation of *logical forms*. (Although Aristotle employs schematic letters for terms, with the aid of which syllogisms can be schematized, he himself has no *doctrine* of logical form.)

One way of approaching Aristotle's problem is to ask, "When one party claims to have refuted another party, is there a principled way of determining whether he has in fact?" Aristotle puts the same question in a slightly different way. How, he asks, can we distinguish *sophistical* refutations from the real thing: and it is to this question that the *Topics*, especially Book VIII, and the *On Sophistical Refutations* are directed.

[2]Aristotle himself had a more complex understanding of dialectic. See Woods and Hansen [2001].

It is a fateful question. It marks a transition in Aristotle's thought from a dominantly dialectical interpretation of arguments to something that can only be thought of as generically *logical*. This last point requires a deft touch. Aristotle did not himself call the new territory into which he was moving by the name of logic. Nor did he apply this name to the mature theory of the *Prior Analytics*.[3] All the same, this is how the tradition that extends from then to the present day thinks of this development in Aristotle's thought, namely, as the invention and development of logic, and there seems to be no good reason to buck the tradition once we see in appropriate detail what it was that Aristotle was up to.

Logic, as we shall now say, was invented by Aristotle to facilitate the solution of a problem in dialectic. It is no more a part of dialectic than a plumber's wrench is a part of a sink. Just as no lover of sinks should disdain wrenches for not being sinks, no lover of dialectic should disdain logic for not being dialectic. Aristotle's logical tool was designed, in the first instance, to mark the distinction between good and bad refutations, themselves clearly dialectical in character. If, as sometimes happens in such cases, the tool took on an interest and achieved a dominance that eclipsed the centrality of what it was a tool for, that doesn't change the fact that for Aristotle logic was introduced as the servant of dialectical argument.[4]

Intuitively speaking, one person's refutation of another's thesis succeeds when the latter's concessions to the former are inconsistent with the latter's own thesis. The heart of the problem is to specify conditions on inconsistency. It is the heart of the problem, but not all of it, as we shall presently see. Aristotle introduces his logical tool with the following words; they constitute a definition of *syllogism*.

> [A] deduction [*syllogismos*] rests on certain statements such that they involve necessarily the assertion of something other than what has been stated, through what has been stated (*On Sophistical Refutations* 1, 165^a 1–3).

As the account develops, it becomes clear that syllogisms are finite sequences of propositions, with the terminal members their conclusions and the preceding members their premisses. Scholars are in agreement that it is true by the definition of 'syllogism' that

1. no such sequence is a syllogism unless its premisses jointly necessitate its conclusion;

2. no such sequence is a syllogism if it contains idle premisses;

[3]The term "logic" seems to have originated with Alexander of Aphrodisias in the third century A.D.

[4]This is disputed. See Hintikka [1987]. For a dissenting opinion, see Woods and Hansen [1997], with a reply by Hintikka [1997].)

3. no such sequence is a syllogism if its conclusion repeats a premiss.

Also of note are two further constraints. In modern systems, the property of validity is *monotonic*. That is, any valid argument stays valid under arbitrary supplementation of its premisses arbitrarily many times. On the other hand, since syllogisms can't tolerate idle premisses, the property of *syllogisity* is non-monotonic. Similarly, whereas a circular argument is trivially valid, syllogisms cannot abide circularity.

Analysis of the structure of syllogisms gives reason to think that additional constraints apply: that syllogisms must be consistently premissed; that syllogisms can have at most one conclusion and at least (and, in some accounts, at most) two premisses; that a logical truth cannot be the conclusion of a syllogism; that the premiss of a syllogism must be relevant to the conclusion; and so on. On the other hand, how closely Aristotle's concept of necessitation resembles the modern notion of logical implication (or entailment) is a matter of contention,[5] but there is quite general agreement that necessitation is primitive in Aristotle's logic (Lear [1985, 8]).

It is well to pause over an important ambiguity. Beyond his interest in disciplining the distinction between good and apparent refutations, Aristotle regarded his logic of syllogisms as the theoretical core of a wholly general theory of argument. But syllogisms are arguments. Does Aristotle think that syllogisms are all there is to arguments? No. Two senses of argument have surfaced. In addition to arguments on the hoof (concerning which there is a process-production distinction), we now have the technical notion which is exhausted by sequences of propositions satisfying the syllogisity conditions and nothing more. Arguments in this second sense are not arguments on the hoof, but Aristotle thinks that they are indispensable to the systematic analysis of arguments on the hoof.

There is another ambiguity which requires flagging. Aristotle's word for syllogism (*syllogismus*) is ambiguous as between *argument* and *reasoning* (or *inference*), as is our word "deduction". There is reason to think that Aristotle intended the ambiguity. Although he would acknowledge that there is a world of difference between *arguments* on the hoof and *inferences* on the hoof, Aristotle appears to think that there were resemblances enough to justify making the theory of syllogisms the theoretical core not only of a general theory of real-life argument, but also of a general theory of real-life (strict) reasoning.

To assist with the task of distinguishing good (or real) refutations from bad (or phony) refutations, Aristotle invokes a dialectical context. As we started to say, in a refutation there are two participants, the proposer, A, of a thesis, T, and the would-be refuter of it, Q. Once having proclaimed T, it is A's role to answer Q's questions honestly and straightforwardly. For his part, Q's questions must be answerable in Yes or No form, and must

[5]See, for example, Woods [2001] for a "proto-modern" answer to this question.

be asked one at a time and clearly. As A produces his answers, Q keeps a record of them. From this record he is free to select any consistent subset as premisses of a syllogism. If A is able to complete a syllogism whose premisses are A's answers and nothing else, and whose conclusion is the contradictory of A's thesis T, then the syllogism in question is a refutation of T.

We see here an interesting mingle of dialectical and logical features. Although the refutation itself is a logical structure, or anyhow as purely logical as any syllogism is, its premisses must have a dialectical origination, just as its conclusion must contradict a dialectically situated thesis. In so saying, it is crucial to keep in mind that a syllogism is itself an entirely non-dialectical entity. It is a plurally-membered sequence of propositions in which the premisses non-superfluously and non-circularly necessitate the conclusion. For these conditions to be met, neither the provenance of the premisses nor the context of their utterance need be taken into consideration. That is, syllogisms are context-free and need not have actually been voiced or advanced by any human arguer. In contrast, a syllogism will not count as a *refutation* of a person's thesis T unless its premisses are drawn exclusively from that person's concessions under press of the other party's questions. We may say, then, that a refutation is a syllogism lodged in a certain dialectical setting.

We begin to see how certain misconceptions arise. If we ask whether the theory of syllogisms is a theory of argument, the answer is straightforwardly affirmative if "argument" is taken in its product sense as a finite sequence of propositions. So construed, a theory of syllogisms is a theory that specifies syllogisms as a proper subset of such sequences; that is, syllogisms would be those very sequences that satisfy certain constraints, including in particular those that we have already listed here. On the present conception, an argument is any finite sequence of propositions, and a syllogism is any such that satisfies the appropriate constraints, e.g., non-superfluous, non-circular necessitation of the terminal proposition by those that precede it.

Even for this rather abstract and bloodless conception of argument, the theory of syllogisms would be at most a subtheory of any would-be theory of argument. A minimal general theory might get away with tracking a basic property of argument-correctness by invoking the necessitation condition. Those satisfying it could be forwarded as truth-preserving sequences and the distinction at hand between the minimally good and the minimally bad could be redeemed as turning on whether the arguments in question preserve the truth of premisses or not. Satisfaction of the necessitation condition is only part of the story of syllogisms. We could, if we liked, reserve the name of argument for something more robust, and proclaim as a condition on such arguments satisfaction of either or both of the conditions of premiss non-superfluousness and non-circularity. Upon doing so, it would be both right and yet hardly pointful to say that the theory of syllogisms was not a theory of argument in all variations of the bloodless conception. It is right

in as much as it is not a theory of argument validity; and it is beside the point in as much as it is never aspired to give such an account.

If "argument" is available to the theorist as a more or less bloodless notion, how much more attractively does it forward itself as an n-person social interaction involving an ordered exchange of speech acts set in a context, and irreducibly influenced by it? These are arguments on the hoof whose description, to say nothing of assessment, requires the application of rules that greatly exceed the reach of logic, even generously construed. Such rules include the whole procedural canon of who speaks when and in what manner, in relation, over all, to what issue. If the theory of syllogisms could not dream of being a complete theory of bloodless argument in every logically relevant variation, how much greater its incompleteness as a theory of argument on the hoof? If a syllogistic logician were to claim comprehensiveness with respect to arguments on the hoof he would be naively wrong, and deservedly the object of a critic's reproof.

Aristotle was as aware of these things. What is interesting about his logic of syllogisms is the extent to which he tries to bring bloodless and on-the-hoof arguments into closer theoretical alignment. In this respect, Aristotle anticipates a boast made by C.I. Lewis about his own logic of "strict" implication, which (he said) may claim

> [as a] primary advantage over any present system ... that its meaning of implication is precisely that of ordinary inference and proof. (Lewis [1912, 531])

It might be questioned how a logic of the implication relation could begin to capture even the rudimentary nuances of proof and of inference. Beyond the minimal claim that Lewis' systems make possible the characterization of deductive or strict validity for proofs and for inferences, it is certainly not true that the logic of strict implication tells us anything more about proof in the ordinary sense or about inference either. Aristotle tries to do better than Lewis, if the anachronism might be forgiven. Aristotle's logic is not a logic of implication. As we have noted, implication is primitive in his writings. Aristotle's syllogistic structures are carefully contrived with certain features of arguments on the hoof clearly in mind. Arguments on the hoof are low-finite exchanges, and the non-monotonicity of syllogisity mimics this feature. In constructing an argument on the hoof for a desired conclusion, the actual reasoner does not make arbitrary searches of arbitrarily large sets of putative premisses. He attempts to economize his effort, and so is drawn to the task of working out premisses relevant to his purposes. Whatever this amounts to in fine, the real-life arguer is disposed to avoid, or at least minimize, the deployment of superfluous premisses, another echo of syllogisms. Nor is the arguer-on-the-hoof satisfied with a deduction whose conclusion merely repeats a premiss. Valid though the deduction is, it is not in any sense a

proof or argument *for* that conclusion, a furtherfact that Aristotle captured in his account of syllogisms.

It could fairly be said that arguments on the hoof imbibe some of the features of ordinary inference or belief-revision. Finite, nonmonotonic, linear, relevant and non-circular[6]: ordinary inferences and syllogisms come to the same thing on some scores, as we have said. To the extent that arguments on the hoof and ordinary inferences also come to the same thing on those scores, it can be seen that Aristotle was able to capture more of argument on the hoof than any theory of implication could imagine itself doing. It is precisely so that Aristotle has the greater claim than Lewis upon inference and proof in their ordinary meanings.

Refutation, again, is a case in point. In Aristotle's account, a refuter of any opponent's position fails in his purpose unless he is able to derive syllogistically the contradictory of that position from his opponent's own concessions, which the refuter is free to use as premisses in the construction of his refuting syllogism. The modern theorist of the implication relation has nothing like this to offer. He can offer only that a refuter must construct a *valid argument* to that same conclusion from those concessions, and for this it suffices to induce the opponent to give inconsistent answers, from which (classically) every validity follows. However we characterize the result of such an exchange, it is stretching things utterly to say that it constitutes a refutation of the particular proposition that chances to be the opponent's position or thesis.

Standard systems of logic are those that track the concepts of logical truth and implication (or, for bloodless arguments, argument validity). As they stand, they will not have much to say about arguments on the hoof, or so it would appear. That alone is some reason for not allowing standard systems to stay all that standard, as witness 20th century developments in relevant logic (Anderson and Belnap [1975] and Anderson *et al.* [1992]). Relevant logic falls outside the plan of the present chapter, but a small observation can be made in passing. In virtually all going systems of relevant logic, the theorist does one of two things, and sometimes both. Either he adds a constraint upon the implication relation so as, among other things, to avert the standard theorem that an implication with inconsistent antecedents is logically true for any proposition as consequent. Or, he adds a constraint upon what to count as a deduction from hypotheses, so that a deduction that fails to use all those hypotheses is not a valid deduction from them. In the first case, he disarms our complaint that a refutation of any thesis could be contrived trivially from any set of inconsistent concessions. In the second case, he is in effect disclaiming the monotonicity of validity. Each time, the further constraint has, or can be given, an Aristotelian motivation. But it is Aristotelian motivation with a difference. The

[6]Also intuitionistic, since a further condition on syllogisms bans multiple conclusions.

relevantist's constraint, whether on implication or on deductions from hypotheses, is meant to redeem the implication relation and the structure of deductions from premisses. That is, the relevantist is bound to say that unconstrained implication isn't implication and that unconstrained deductions from premisses aren't valid deductions. Aristotle need not say this and almost certainly doesn't believe it. By his lights, unconstrained deductions from premisses are valid. What they aren't is *syllogisms*. Readers may wish to ponder whether we have here a difference that makes a difference.[7]

3 DEDUCTIVE LOGIC

Though originating with the Stoics, propositional logic emerged in modern dress in the writings of George Boole (1815–1864). Boole invented a 'calculus' in which infinitely many argument forms of arbitrary complexity are valid (Boole [1847, 1854]). Other precursors are Augustus De Morgan (1806–1871) and Charles S. Peirce (1839–1914). Aristotle's syllogisms are finite sequences of propositions. "Proposition" is a technical term for Aristotle. In modern terms, an Aristotelian proposition is a statement whose only logical particles are quantifier expressions and expressions for something like term complementation. Thus, for Aristotle neither

Socrates is a man or the cat is on the mat

nor

If Harry punched Bill then Mary will be angry

can be constituents of syllogisms.

In a **propositional logic**, there are two main departures from what modern readers will now recognize as Aristotle's quantifier-negation-copula fragment of some "more capacious logic", as one commentator has said — something approximating to the modern logic of quantifiers, or the calculus of predicates. In propositional logic, quantifiers, terms and the copula are suppressed; and connectives other than negation are admitted. (Barnes [1994, vxi]) Though quantifiers, terms and the copula are suppressed they are suppressed in different ways. For reasons that will later become clear, no sentence on the hoof containing a quantifier has a representation in propositional logic. On the other hand, sentences containing terms and the copula can be represented in propositional logic, but their mode of representation is such that neither their terms nor the copula is reflected in the representation.

The representation of "Mary is angry" is by way of an *atomic sentence* of propositional logic. Any atomic sentence to which "Mary is angry" might be assigned is one of up to infinitely many propositional letters,

[7]Concerning which see Woods [2002].

$p_1, p_2, \ldots, p_n \ldots$. For ease of exposition, it is useful to supplement the set of atomic sentences with *p, q, r, s, t*, etc. The minimal constraint on sentences on the hoof that are eligible for representation as atomic sentences of propositional logic is that they must be connective-free; that is, they must not contain expressions such as "it is not the case that" (or negative prefixes such as "im" or "un"), "or", "and", "if. . . then", and the like. As will become apparent, this is a far from satisfactory characterization of the sorts of sentence on the hoof that have atomic representations in propositional logic. But the main idea is clear enough to be getting on with. We turn now to an elementary system of propositional logic, the *propositional calculus* **PC**.

PC is a formal grammar with a set of interpretations. The grammar stipulates a *vocabulary*, and then builds up from it compositionally a set of *sentences* of **PC** and a set of *sequences* of **PC**-sentences. With the grammar fixed, the set of interpretation and function relations is displayed in such a way as to make possible both the characterization and effective recognition of a stock of target properties. For **PC** sentences these include the properties of tautologousness, contingency and self-contradiction. For **PC**-sequences, target properties include argument-validity, and deducibility; and, for sets of **PC**-sentences; consistency. Beyond these are properties of properties, e.g., the decidability of validity and the compactness of deducibility (see below). Then come properties of sets of **PC**-expressions, such as functional completeness, and system-wide properties, such as completeness and soundness (see below). In our exposition here we keep to an elementary account. Our purpose is to expose enough of the logical structure of **PC** to make its contribution to a theory of argument and of inference a judgeable matter.

The convention by which connective-free elementary sentences on the hoof are represented in propositional logic by sentence-letters gives rise to an elementary notion of sentential (or propositional) *form*. We may say that the simple atomic sentences $p, q, r, \ldots, p_1, p_2, \ldots$ reflect the forms of the natural language sentences whose representation they are. This is accomplished in two ways. First, by their atomicity the p_i indicate that what they represent have the form of simple unquantified sentences. Second, the fact that the p_i themselves have no internal grammatical structure indicates that the internal grammatical structures of the simple unquantified sentences which they represent are deemed for the purposes at hand to be logically inert. Thus the p_i represent the form *but not the content* of quantifier-free simple sentences of English.

In addition to the atomic sentences, the vocabulary of **PC** includes the *connectives*, 'not', 'and', 'or', 'if. . . then' and 'if and only if', and the *punctuators*, left parenthesis '(' and right parenthesis ')', and the sequence markers, '$\langle$' and '$\rangle$'. It is well to note that the Greek upper case symbols Φ, Ψ are not expressions of the language of **PC**. They belong to English and, as here used, are variables ranging over sentences or sentence-forms of **PC**. English

is the language in which we are currently describing the language of **PC**. So English is, among other things, language about language. We acquiesce in an etymologically questionable convention and shall say that English is the *metalanguage* for **PC**, whose own (formal) language is the indexformatin rules*object language* of **PC**. Accordingly, in our Formation Rules Φ and Ψ are metalinguistic variables that range over expressions constructed out of **PC**'s vocabulary. In those same rules, the connectives occur autonymously, that is, as names of themselves. It is also desirable to have a policy for the mix of names of connectives and metalinguistic variables, Φ, Ψ, χ, etc. One such is Quine's method of quasi-quotation. We will explain this a short paragraph hence.

The *elementary grammar* of **PC** is a set of Formation Rules which when applied to the vocabulary produces an effectively recognizable set of **PC**-sentences. The Formation Rules provide as follows: that an atomic sentence of **PC** is a sentence of **PC**; that if Φ is a sentence of **PC** so too is $\ulcorner$not-$\Phi\urcorner$; that if Φ, Ψ are sentences of **PC** so too are $\ulcorner\Phi$ and $\Psi\urcorner$, $\ulcorner\Phi$ or $\Psi\urcorner$, $\ulcorner$if Φ then $\Psi\urcorner$, $\ulcorner\Phi$ if and only if $\Psi\urcorner$, and that nothing else is.

How do we interpret expressions such as

$\ulcorner$not-$\Phi\urcorner$
$\ulcorner\Phi$ or $\Psi\urcorner$

and so on? What, in particular, is the role of the symbols '$\ulcorner$' and '$\urcorner$', or Quine-corners, as they are also called? (Quine [1950]) The variables 'Φ' and 'Ψ' are metalinguistic variables. They range over formulas of **PC** but are not themselves **PC**-formulas. Consider the Formation Rule

R_1 If Φ is an atomic sentence of **PC**, then so too is $\ulcorner$not-$\Phi\urcorner$.

The corner-quotes function selectively. They quote the contained **PC**-expression but do *not* quote the metalinguistic variables. Thus R_1 says that for any entity in the range of the variable, if it is an atomic sentence of **PC** (no quotation of any such formula, so far), then the result of prefixing to that same sentence (no quotation, yet) an occurrence of the connective '$\neg$' (quotation!) is likewise a sentence. Similarly for

R_2 If Φ and Ψ are **PC**-sentences, so too is $\ulcorner\Phi$ or $\Psi\urcorner$.

Here the rule says that for any entity in the range of the left-variable and any entity in the range of the right-variable are such that if they are **PC**-sentences the same is true of the result of inserting an occurrence of 'or' between those two sentences in that order.

In the *extended* or *full* grammar of **PC**, additional rules come into force. One is the *sequence-rule*: If $\Phi, \ldots, \Phi$ are **PC**-sentences, then $\langle\Phi_1, \ldots, \Phi_n\rangle$ is a **PC**-sequence (of sentences). The others are Transformation Rules which are rules for constructing **PC**-sequences in such a way that some target

property (e.g. the property of being a *theorem*) is transmitted from initial member(s) of the sequence to its terminal member. Transformation rules are discussed in a later section.

The grammar of **PC**, whether in its elementary or full dress, extends the notion of logical form in rather natural ways. The full notion of sentential form in **PC** is captured recursively by saying that when Φ, Ψ are sentence-forms, atomic or molecular, so too are : $\ulcorner$not-$\Phi\urcorner$, $\ulcorner$not-$\Psi\urcorner$, $\ulcorner\Phi$ and $\Psi\urcorner$, $\ulcorner\Phi$ or $\Psi\urcorner$, $\ulcorner$if Φ then $\Psi\urcorner$, $\ulcorner\Phi$ if and only if $\Psi\urcorner$, and nothing else is. Thus in **PC** a sentence is a sentential form.

Sequence-forms, in turn, are finite sequences of sentence-forms. If Φ, $\ldots \Phi_n$ are sentential-forms of **PC**, then $\langle \Phi_1, \ldots, \Phi_n \rangle$ is a sequence-form of **PC**, and nothing else is. If $\langle \Phi_1, \ldots, \Phi_n \rangle$ is a sequence-form generated by the Transformation Rules of **PC** and if Φ_1 has the theoremhood property, then $\langle \Phi_1, \ldots, \Phi_n \rangle$ is a *proof-form* (or for lexical relief, a *formal proof*) of sentence-form Φ_n from sentence-form Φ_1.

Axiomatic approaches to propositional logic are wholly syntactic. (Frege [1879], Hilbert [1923]) There are also variations of which the best-known is the *natural deduction* approach. It has advantages that axiom systems lack. One is that it operates without the necessity of having to specify (and justify) a distinguished set of axioms. Another is that it permits the testing of arguments for validity in ways that are said better to resemble patterns of reasoning on the hoof (e.g. Kalish and Montague [1964]). Modern natural deduction approaches originated independently with Gentzen [1934] and Jáskowski [1934] and exist in several forms: e.g., Mates [1965], Fitch [1952], Kalish and Montague [1964], Lemmon [1978], Prawitz [1965], Quine [1950], Suppes [1957], Thomason [1970], Tennant [1978] and van Dalen [1980]. A third approach is that of the sequent calculus (Gentzen [1934]. We shall here expound the truth table method.[8] An axiomatic approach will lightly be sketched in a subsequent section.

PC is a sequence $\langle$*Grammar, Set of Interpretations*$\rangle$ or $\langle \mathbf{G}, \mathbf{I} \rangle$ for short. Here '**G**' is ambiguous as between the elementary grammar and the full. Until further notice, we shall confine our remarks to the elementary grammar. **I** is a set of interpretations i of **G**. An interpretation of **G** is a function from the atomic sentences of **G** to a pair-set of objects called *truth-values*. By a natural convention, the truth-values of **PC** are identified as *truth* (T) and *falsity* (F). To facilitate this requirement of joint exhaustiveness it is necessary to read 'false' as 'not true'. In strictness, any pair of objects $\{X, Y\}$ will do as truth-values in **PC** provided that every sentence-form is assigned either X or Y and none is both.

In giving its interpretation of **G**, each $i \in I$ assigns to each atomic sentence-form of **PC** one or other of T, F. The assignments to the total

[8]Alternatives to the truth table approach include various versions of the *semantic tableaux* method, originated by Beth [1955] and Hintikka [1955], as well as the *fell swoop* technique developed by Quine [1950].

set of the atomic sentences are 2^n-many, where n is the number of atomic sentences of **PC**. Thus **I** provides an exhaustive set of distinct total truth-value assignments to the set of atomic sentences of **PC**.

We now introduce the concept of a valuation. A valuation v is an extension of an intepretation which is defined on all sentences of **PC**. Valuations in **PC** are subject to the following conditions.

Where Φ is a sentence-form of **PC**, it is customary to write '$v(\Phi) = \pm T$', which is read as "the truth-value of Φ is T" (or F as the case may be).

Then, if $v(\Phi) = T, v(\text{not-}\Phi) = F$, and if $v(\Phi) = F, v(\text{not-}\Phi) = T$. That is, Φ is true if and only if $\ulcorner$not-$\Phi\urcorner$ is false. $\ulcorner\Phi$ and $\Psi\urcorner$ is true if and only if Φ is true and Ψ is true. $\ulcorner\Phi$ or $\Psi\urcorner$ is false when both Φ and Ψ are false and is true otherwise. $\ulcorner$If Ψ then $\Psi\urcorner$ is true except where Φ is true and Ψ is false. $\ulcorner\Phi$ if and only if $\Psi\urcorner$ is true except where Φ and Ψ have different truth values. Schematically,

- $v(\Phi) = T$ if and only if $v(\text{not-}\Phi) = F$
- $v(\Phi \text{ and } \Psi) = T$ if and only if $v(\Phi) = T$ and $v(\Phi) = T$
- $v(\Phi \text{ or } \Psi) = T$ if and only if at least one of the following propositions holds: $v(\Phi) = T$; $v(\Psi) = T$
- $v(\text{if } \Phi \text{ then } \Psi) = T$ except where $v(\Phi) = T$ and $v(\Psi) = F$.
- $v(\Phi$ if and only if $v(\Phi) = v(\Psi)$

From these interpretation-rules it is easy to see that non-atomic or molecular sentences of **PC** are true or false depending entirely upon the truth or falsity of their atomic constituents. In every case the truth-value of a **PC**-sentence is uniquely determined by, or is a function of, the truth values of its atomic sentences and hence, given an interpretation $i \in I$ there is exactly one valuation v that extends i ($i \subseteq v$). In view of this, the connectives of **PC** have come to be called "truth functional" connectives, and **PC** itself is called *truth functional logic* or the *logic of truth functions*.[9]

With the truth-value rules ready to hand, a number of target properties of **PC** become explicitly definable. If, for example, Φ is a sentence-form of **PC** then Φ is a logical truth, or *tautology*, if and only if Φ is true for each truth-value assignment to all its atomic constituents. Φ is *contingent* if and only if it is true for at least one such atomic-assignment and is false for at least one other such assignment. Φ is a *self-contradictio*n if and only if it is false for all atomic assignments.

[9]**PC** is functionally complete. That is, any truth function of n-tuples of **PC**-sentences is definable in terms of the present five. These in turn reduce to one or other of the pairs {'not', ' and '}, {'not', 'or'} or {'not', 'if...then'}. We shall demonstrate this claim in due course.

Logical truth in **PC** is a fruitful notion.[10] Not only does it explicitly characterize the intuitive notion of a proposition which could not in any sense be false, but it makes possible (what Aristotle's logic lacked) an explicit definition of *argument-validity*. We here re-introduce our bloodless conception of argument. Let us say that any finite sequence of **PC**-sentence forms is an *argument-form*. Thus $< \langle \Phi_1, \ldots, \Phi_n \rangle$ is the argument-form in which Φ_n is the *conclusion* and the members preceding it are *premisses*. For every argument-form in **PC** there is a conditional (or "if...then") sentence-form which corresponds to it uniquely. If $\langle \{\Phi_1, \ldots, \Phi_{n-1}\}, \Phi_n \rangle$ is argument-form then its *corresponding conditional* is (If $\Phi_1, \wedge \ldots, \wedge \Phi_{n-1}$, then Φ_n). Then $\langle \Phi_1, \ldots, \Phi_n \rangle$ is *valid* and if and only if (If $\Phi_1, \wedge \ldots, \wedge \Phi_{n-1}$ then Φ_n) is a tautology.

It is always possible to determine whether a **PC**-sentence-form is a tautology, and possible, as well, to make these determinisms infallibly, mechanically and in finite time. This being so, the property of tautologousness is *decidable* or *effectively recognizable* in **PC**. Since an argument-form is valid in **PC** just when its corresponding conditional is a tautology, validity too is decidable in **PC**.

The best-known procedure for checking for the tautologousness property in **PC** is the *truth-table method*. We will here illustrate the method intuitively and informally with two examples. Consider the two **PC** sentences

1. If (p_1 and (if p_1 then p_2)), then p_2

and

2. p_4 if and only if (p_5 and p_{10}).

Beginning with (1), we construct its truth-table as follows. We set out a top row of the table. The top row is divided into a left part and a right. In the left part, the atomic sentences of (1) are set out, usually in the order of their occurrence as in (1). In the right part, (1) itself is displayed. Thus

Left			**Right**
Top row	p_1	p_2	If (p_1 and (if p_1 then p_2)), then p_2

We must next calculate how many additional rows are required. This done by the 2^n-rule, where n is the number of atomic sentences in (1). In this case, there are two; so the number of rows is 2^2 or four. Whereupon,

[10]A hefty abundance of tautologies can be found in Kalish and Montague [1964, 80–84].

Left		**Right**
Top row	p_1 p_2	**If** (p_1 **and** (**if** p_1 **then** p_2)), **then** p_2
Row 1		
Row 2		
Row 3		
Row 4		

Next is required a mechanical method for assigning to the atomic p_1 and p_2 each of their distinct truth-value assignments. One way of doing this is as follows. In the *column* under the left most atom in Left (i.e., under p_1), enter T and alternate with F every once in each succeeding row. Then, in the column under p_2, enter T in the first two rows, and vary equally with F every two times (i.e., in every succeeding two rows). This gives

Left		**Right**
Top row	p_1 p_2	**If** (p_1 **and** (**if** p_1 **then** p_2)), **then** p_2
Row 1	T T	
Row 2	F T	
Row 3	T F	
Row 4	F F	

The method of atomic assignments easily generalizes. In the left most column begin with T and vary with F every once until 2^{n} entries have been made in that column. For any other column, enter half the number of of Ts entered in its neighbouring column to the left and vary with Fs at twice the rate of the variations of Ts and Fs in the neighbouring column.

With the atomic assignments in place, and using our definitions of the connectives 'and' and 'if ... then', we assign row by row a truth value to constituent sentences of (1). This is done by way of the *length rule*. We say that the length of a **PC**-sentence is the number of symbols comprising it (not counting parentheses or subscripts). In (1) the atomic sentences p_1 and p_2 are of *least length*. The length rule provides that assignments in the rows in Right start with sentences of least length, the next-least length, then next-next-least length, and so on until one has a truth-value for all of (1) in the row in question. Where a sentence has more than one constituent sentence of any given length, the length rule stipulates that they be evaluated in the order of their appearance, left to right, in the complete sentence whose truth-table is under construction. In our example, the length rule gives us a complete table:

Left			**Right**
Top row	p_1	p_2	**If** (p_1 **and** (**if** p_1 **then** p_2)), **then** p_2
Row 1	T	T	T T T T T T T
Row 2	F	T	F F F T T T T
Row 3	T	F	T F T F F T F
Row 4	F	F	F F F T F T F

The numbers at the feet of the columns indicate the order of construction. Statement (1) is true in every row of its truth-table. In each row, truth-value assignments are made to the atomic constituents of (1) and, given the definitions of the connectives of (1), an assignment is made to it. We may say then that each row gives a *interpretation* of the atomic constituents. Since in each row, (1) is true given the atomic assignments in that row, we see that (1) is true for all it atomic assignments. Since assignments to atomic sentences that do not occur in (1) make no difference, this gives an alternative characterization of tautologousness: Φ is a tautology of **PC** if and only if it is true for every valuation on its atomic consistents.

Given that our rules of truth-table construction provide that every **PC**-formula has a unique and effectively recognizable truth-table, we may also say that the truth-table method is a *decision-procedure* for tautologousness in **PC**, hence here too for argument-validity in **PC**. Our second example is more briefly dealt with.

Left				**Right**
Top row	p_4	p_5	p_{10}	p_4 **if and only if** (p_5 **and** p_{10})
Row 1	T	T	T	T T T T T
Row 2	F	T	T	F F T T T
Row 3	T	F	T	T F F F T
Row 4	F	F	T	F T F F T
Row 5	T	T	F	T F T F F
Row 6	F	T	F	F T T F F
Row 7	T	F	F	T F F F F
Row 8	F	F	F	F T F F F

Whereas sentence-form (1) had just two atomic consistents, (2) has three. So our 2^n-rule requires that its truth-table have 2^3 or eight rows. In each column, the rule for the arraying of Ts and Fs works as before, except that in each column the rule must operate long enough to generate $2^2 = 8$ entries. Another difference occasioned by the increase from two to three atomic consistents is that under the third, p_{10}, we must enter four Ts (half the number entered under p_5) and vary equally with Fs every four times (twice the variation rate under p_5). If we had had a fourth atom, occurring fourth in the sentence whose truth-table is in question, say p_{17}, then the

rule at hand would have displayed Ts and Fs, for a total of sixteen rows, in the following pattern for the fourth constituent:

p_{17}
T
T
T
T
T
T
T
T
F
F
F
F
F
F
F
F

With rows understood to give interpretation of atomic constituents, we see that (2) is true on some interpretations and false on others; i.e., that (2) is contingent and not a tautology. It is also apparent that "one-half" of the *biconditional* sentence that (2) expresses (denoted by the connective 'if and only if'), namely,

(2*) If p_4 then (p_5 and p_{10}),

is also contingent. Hence the argument-form to which it uniquely corresponds, $\langle$If p_4 then (p_5 and p_{10})$\rangle$, is not valid, but *invalid*. (The same holds for the other half.)

We have been describing the operation of our set of interpretation functions i on the elementary grammar **G** of **PC**. Because every **PC**-sentence has a unique truth-table (cf. Kleene [1952, 21ff]), **I** permits the effective recognition for arbitrary **G**-structures of the following properties: for sentences: tautologousness, contingency and self-contradiction: for sequences, validity and invalidity of argument-forms; and for (finite) sets of sentences: consistency and inconsistency. (For let $\sum$ be any set of **PC**-sentences. Then $\sum$ is *consistent* if and only if for at least one valuation on the atomic constituents of the member sentences of $\sum$ all member sentences are true). So, again, **PC** is *decidable* with respect to its target properties π_i. For any such π_i it is possible to determine infallibly, mechanically and in finitely many operations whether a PC-structure has π_i.

Concerning argument validity (or entailment), PC also has the Deduction Property: if $\langle \Phi_1, \ldots \Phi_{n-1}, \Phi_n, \Psi \rangle$ is valid then $\langle \Phi_1, \ldots \Phi_{n-1}$, if Φ_n then Ψ is valid. This is a form of the Deduction Metatheorem. It is an aid, if only it were needed, in recognizing valid entailments in **PC**. **PC** also answers to Interpolation Metatheorem: If $\langle \Phi, \Psi \rangle$ is valid, there exists a formula χ such that $\langle \Phi, \chi \rangle$ is valid and $\langle \chi, \Psi \rangle$ is valid, where every atomic sentence in χ likewise occurs in Φ and Ψ)(or $\chi = p$" and not p" or $\chi =$ 'p, or not p')[11] (Craig [1957]).

We close this section with a brief anticipation of something that will deal we will deal with more fully when we discuss formalization of natural language constructions, in section 1.6. All that we say for now is that for technical reasons we now drop from **PC** the connectives, 'not', 'and', 'or', 'if...then' and 'iff and only if', and replace them in order with the autonymous expressions $\neg$, $\wedge$, $\vee$, $\supset$, and $\equiv$. We also stipulate that in all strictness our truth table definitions of the connectives are definitions of this second bunch. The connection with their natural language counterparts is discussed two sections hence.

4 AN AXIOMATIZATION OF PC

In our treatment of **PC** so far, the target properties of **PC** are delivered by its syntax (e.g., the property of sentencehood) or by its truth-tabular provisions (e.g., the property of logical truth or tautologousness). In the present section we shall briefly indicate a way of generating in a wholly syntactic way exact counterparts of the logical properties generated by **PC**'s truth-tabular apparatus. Properties generated in this latter way are said to be semantic properties.

It is provable that there is an exact concurrence between these semantic properties and their merely syntactic counterparts in the axiomatization of **PC**. In A(**PC**) the counterpart of the property of tautologousness is the property of theoremhood. The theorems of A(**PC**) are precisely those **PC**-sentences derivable from the axioms. By the Completeness Metatheorem, every tautology is a theorem (or is provable). By the Soundness Metatheorem, every **PC**-provable sentence (or theorem) is a tautology in **PC**. We will demonstrate the Soundness Metatheorem below.

In A(**PC**) truth values have no role. Target properties are specified by adding to the existing grammatical rules of **PC** some additional grammatical rules and definitions. One of the important features of A(**PC**) is its capacity to generate logical properties just by manipulating grammatical forms.

A(**PC**) arises from the grammar of **PC** in the following way. We introduce first a finite set of axioms. Axioms are understood in a technical way.

[11] Or $\chi = \perp$ or $\chi = T$ where $\perp =$ *Falsum* (or "bottom") and $T =$ *verum* (or "top").

They are introduced by stipulation, and they do *not* carry this ordinary semantic meaning of being self-evidently true. They are peices of *syntax*.

*Axioms of A(**PC**)*

1. $p \supset (q \supset p)$

2. $p \supset (q \supset r) \supset ((p \supset q) \supset p \supset r))$

3. $(\neg p \supset \neg q) \supset (q \supset p)$

The next step is to specify a finite set of *transformation rules*. Transformation rules (sometimes and misleadingly called "rules of inference") when applied to axioms produce *theorems*; and when applied to theorems produce more theorems, and so on. Thus the set of theorems of $A(\mathbf{PC})$ is said to be the deductive closure of the $A(\mathbf{PC})$-axioms under the transformation rules. Here, too, the term "theorem" is a technical one. It carries none of the ordinary semantic meaning of a "demonstrated truth". Theorems, like axioms, are peices of syntax.

Transformation Rules of $A(\mathbf{PC})$

R1. From Φ and $\ulcorner(\Phi \supset \Psi)\urcorner$, it is permissible to derive Ψ.
(*Modus Ponens*)

R2. From Φ, it is permissible to derive the result of substituting a sentence Ψ for an atomic sentence in Φ, in all its occurrences.
(*Substitution*)

We are now able to characterize *proofs* in $A(\mathbf{PC})$.

Proof A proof in $A(\mathbf{PC})$ is any finite sequence of **PC**-sentences each member of which is either an axiom or can be derived from preceding members of the sequence by applying a transformation rule.

Now for theorems.

Theorem A theorem of $A(\mathbf{PC})$ is the terminal member of any $A(\mathbf{PC})$-proof.

We are here giving only the briefest glimpse of the basics of an axiomatic system. In the interests of space, we shall not take the time to study techniques for the construction of proofs. There is a plethora of accessible elementary accounts of this. (Woods [2000]) However, as mentioned, we shall demonstrate that $A(\mathbf{PC})$ satisfies the Soundness Metatheorem (see just below).

5 TWO METATHEOREMS

We now give the reader a glimpse of the metatheoretical potency of **PC**. A theorem of **PC** is a **PC**-sentence that is either a **PC** axiom or is in the deductive closure of the **PC**-axioms. On the other hand, a **PC**-metatheorem is a sentence of English which states some deomonstrable fact *about* **PC** (not *in* it), where the demonstration of that fact is likewise constructed in English (or whatever home language the theorist is using). We shall here demonstrate the Functional Completeness Metatheorem and the Soundness Metatheorem. The first reports of a *truth-functional* (or semantic) fact; the second reports a combined fact about semantics and syntax.

Functional Completeness
Any truth functional connective of abitrarily many places can be replaced without relevant loss by either of the pairs $\{\neg, \wedge\}$, $\{\neg, \vee\}$.

Metaproof By the construction rules of truth tables, every truth functional connective has a defining truth table. Since **PC**-sentences are of finite length, every **PC**-truth table is finite.

> Let K be an abitrarily selected truth functional connective of arbitrarily many places (dyadic, triadic, ..., n-adic, ...). Let M be the defining truth table for K. Let K-Φ be the formula of the defining truth table.

There are three cases to consider.

Case one. There is no row of M in which Φ comes out true. In that case, Φ is false for every valuation, and so has the same truth conditions as $\ulcorner(\Psi \wedge \neg\Psi)\urcorner$. Thus any K-sentence is equivalent to a $\{\neg, \wedge\}$-sentence. K is eliminable in favour of $\{\neg, \wedge\}$.

Case two. There is no row of M in which Φ comes out false. In that case, Φ is true for every valuation, and so has the same truth conditions of $\ulcorner(\Psi \vee \neg\Psi)\urcorner$. Thus any K-sentence is equivalent to a $\{\neg, \vee\}$-sentence. K is eliminable in favour of $\{\neg, \vee\}$.

Case three. For some rows of M Φ comes out true, and for others it comes out false. Consider now any row in which Φ comes out true. In that row the atoms of Φ will be either true or false. Define a *literal* of **PC** as any atom or its negation. Now construct conjunctions of the literals in that row according to the rule that if an atom is true in that row then the literal is that atom, and if an atom is false in that row then the literal is its negation. Repeat the same construction of conjunctions of literals in every other row of M in which Φ comes out true. Let C_n be any such conjunction, with the subscript n denoting the number of rows in question.
The *characteristic formula CF* of M is the disjunction of the C_i, that is, the disjunction of the conjunction of literals in each row in which Φ comes

out true. The characteristic formula thus gives the truth conditions for the formula Φ of the defining matrix M. But Φ is any arbitrarily selected K-sentence, and CF is truth conditionally equivalent to it. Since CF requires no connectives other than $\{\neg, \wedge, \vee\}$, then K is eliminable in favour of that trio.

Having shown that any truth-functional K is eliminable in favour of $\{\neg, \wedge, \vee\}$, it suffices to show that $\{\neg, \wedge, \vee\}$ is eliminable in favour of either $\{\neg, \wedge\}$ or $\{\neg, \vee\}$. This is done by truth values that demonstrate the De Morgan equivalences:

1. $(\Phi \wedge \Psi) \longleftrightarrow \neg(\neg\Phi \vee \neg\Psi)$

2. $(\Phi \vee \Psi) \longleftrightarrow \neg(\neg\Phi \wedge \neg\Psi)$

We also have it that the pair $\{\neg, \supset\}$ suffices for complete truth functional expressive power. By the so-called *implication law*, any sentence $\ulcorner(\Phi \supset \Psi)\urcorner$ is truth conditionally equivalent to $\ulcorner(\neg\Phi \vee \Psi)\urcorner$ which, in turn, by De Morgan's Law is truth conditionally equivalent to $\ulcorner\neg(\Phi \wedge \neg\Psi)\urcorner$.

To demonstrate soundness, we introduce some needed definitions.

Satisfiability A set of **PC**-sentences is satisfiable iff there is a valuation such that all its member-sentences are true.

(Note, that A(**PC**) structures can *have* semantic properties. But they cannot be provided or demonstrated by A(**PC**) itself. A(**PC**) and **PC** have exactly the same *sentences*, and we already have it that every **PC**-sentence is either true or false and, accordingly, that every **PC**-sentences has a valuation.)

Logical truth (or sentence-validity)
A **PC**-sentence is a logical truth (or a valid sentence) iff it is satisfied by every valuation.

Finally, we will also need the principle of strong mathematical induction. It is customary to represent this principle as an argument whose conclusion is derived from two premisses, one called the *basis*, and the other the *induction step*, as follows

Mathematical Induction

1. The number 0 has property P (Basis)

2. If all predecessors of any natural number n have P, does too does n have P. (Induction step)

3. Therefore, all natural numbers have P. (Conclusion)

Now for the Soundness Theorem.

Soundness Every theorem of $A(\mathbf{PC})$ is valid.
Metaproof It suffices to show that the $A(\mathbf{PC})$ axioms are valid (logically true; tautologous) and that validity is preserved by the transformation rules.

It is easy to verify by truth tables that the $A(\mathbf{PC})$-axioms are valid.

We now show that *modus ponens* preserves validity. We assume the contrary, that Φ and $\ulcorner(\Phi \supset \Psi)\urcorner$ are valid and that Ψ is not valid. Since Ψ occurs in $\ulcorner(\Phi \supset \Psi)\urcorner$, there is a valuation v such that Φ and $\ulcorner(\Phi \supset \Psi)\urcorner$ come out true but Ψ comes out false. But if Φ is true and Ψ is false, then $\ulcorner(\Phi \supset \Psi)\urcorner$ is false, contrary to the assumption. And if $\ulcorner(\Phi \supset \Psi)\urcorner$ is true and Ψ is false, then Φ is false, contrary to the assumption. Thus *modus ponens* is validity-preserving.

We now show that the substitution rule preserves validity. To do requires that we demonstrate the following lemma (or preliminary metatheorem).

Lemma Let Φ_Ψ arise by substituting Ψ for atomic sentence α in Φ_α. Let $\alpha_1, \ldots, \alpha_n$ be a non-repeating array of atoms in which are included α and any atom occurring in Φ_α or Ψ. Finally, let v be a valuation (subsuming an interpretation i) on $\alpha_1, \ldots, \alpha_n$. Then if v provides identically for Ψ and α, then v also provides identically for Φ_Ψ and Φ_α.

Metaproof of the Lemma

1. Suppose that α fails to have an occurrence in Φ_α. Then Φ_Ψ is the same sentence as Φ_α, and the lemma is satisfied trivially.

2. Suppose that α does have an occurrence in Φ_α. If Φ_α is connective-free, then Φ_α is the same sentence as α. So it is obvious that the lemma is satisfied.

3. Suppose that the lemma holds for any sentence Φ'_α with k or fewer occurrences of connectives such that α occurs in Φ'_α. Let Φ_α contain k+1 occurrences of connectives and at least one occurrence of α. Then, by the functional completness metatheorem and the rule of implication, it suffices to put it that Φ_α is a sentence of the form $\ulcorner\neg\chi_\alpha\urcorner$ or $\ulcorner\Omega_\alpha \supset \Upsilon_\alpha\urcorner$.

Consider the options. Suppose Φ_α is $\ulcorner\neg\chi_\alpha\urcorner$. Since $\neg$ is an occurrence of a connective greater than k, and the lemma holds for k-occurrences, then the lemma holds for χ_α. But χ_α just is Φ_α, since Φ_Ψ is the same sentence as $\ulcorner\neg\chi_\alpha\urcorner$.

Suppose next that Φ_α is $\ulcorner\Omega_\alpha \supset \Upsilon_\alpha\urcorner$. Then the lemma holds for Ω_α or it contains one or more occurrences of connectives fewer than k+1. The same reasoning applies to Υ_α. So Υ_α satisfies the lemma.

It follows then, that the lemma holds for $\ulcorner\Omega_\alpha \supset \lambda_\alpha\urcorner$, since $\ulcorner\Omega_\alpha \supset \lambda_\alpha\urcorner$ is the same sentence as Φ_α. This is so because Φ_Ψ is the same sentence as $\ulcorner\Omega^\Psi \supset \lambda^\Psi\urcorner$.

We have it, now, by strong mathematical induction that the lemma is satisfied by any sentence Φ_α in which α has an occurrence. If it does not, the lemma holds trivially, as we have seen.

This completes the metaproof of the lemma.

The assertion that substitution preserves validity now easily follows. The Soundness Metatheorem therefore holds.

Not every **PC**-theorist is interested in modelling natural language argument and inference in his logic. But many are, and have been.

In the form in which we have developed it so far, it is possible to see the *general* rationale of **PC** as not much different from that of Aristotle in his own formal theory of syllogisms, as set out in the *Prior Analytics*. Like Aristotle, the **PC**-logician is sometimes interested in capturing a class of arguments on the hoof. Aristotle was interested in arguments that are syllogisms. The PC-logician has a different goal, at once more ambitious and less than Aristotle's. Unlike Aristotle, the **PC**-logician is content with capturing arguments that are valid, many of which would fail to be syllogisms. In this, his programme is less ambitious than Aristotle's, which aimed at specifying a class of arguments having properties other than validity. As against this, the **PC**- logician is also more ambitious. For, unlike Aristotle, he can offer an explicit definition of validity.

It will not have escaped the reader's notice that, although it is not standard a theoretical target for it, **PC** is nevertheless capable of generating a conception of valid argument which closely resembles Aristotle's syllogism. These we might fairly call **PC**-syllogisms, and we could say that a valid **PC**-argument-form $\langle\{\Phi_1,\ldots,\Phi_n\},\Psi\rangle$ is a **PC**-syllogism if and only if:

(a) No proper subargument of $\langle\{\Phi_1,\ldots,\Phi_n\},\Psi\rangle$ is valid.

(b) In $\langle\{\Phi_1,\ldots,\Phi_n\},\Psi\rangle$, no $\Phi_i = \Psi$; $(i \leq n)$.

(c) $\{\Phi_1,\ldots,\Phi_n\}$ is consistent.

Note that $\langle\{\chi_1,\ldots,\chi_m\}\Psi\rangle$ is a *subargument* of $\langle\{\Phi_1,\ldots,\Phi_n\},\Psi\rangle$ iff $\{\chi_1,\ldots,\chi_m\} \subseteq \{\Phi_1,\ldots,\Phi_n\}$ and a *proper* subargument if $\subseteq$ is replaced by $\subset$. **PC** certainly has the resources to determine effectively for any valid **PC**-sequence whether it satisfies conditions (a) to (c). So the property of being a **PC**-syllogism is decidable in **PC**. And if, as was said in an early section of this chapter, the concept of syllogism adumbrates a more full-blooded concept of argument than that conveyed by the concept of merely valid argument, it would seem that **PC** has the resources to honour the structure of arguments on the hoof approximately as well as the theory of syllogisms (though, of course, they will be a different class of arguments

on the hoof, given the essential differences between the vocabularies of the respective theories). Then, too, if as has been suggested the concept of syllogism is one that brings the structure of (deductive) *inference* into closer alignment than could be brought off by the concept of a merely valid argument, it would seem that **PC** is as well positioned to accommodate a class of deductive inferences as is the theory of syllogisms, though here too, it is a different class of deductions.

It might be supposed that since **PC** and the syllogistic do approximately as well in capturing the structure of good argument and inference in their respective domains of application, they also do approximately as badly. For consider condition (a) on **PC** syllogisms. It provides that no **PC**-argument is a syllogism if its conclusion is a necessary truth (Thom [1981, 22], and Woods [2001, chapter 7]). Given **PC**'s definition of validity, any valid **PC**-argument whose conclusion is a necessary truth is one whose conclusion is also validly derived from any proper subset of the argument's premisses, including the empty set of premisses. Whereupon no **PC**-proof of a necessary truth can be a **PC**-syllogism. This is odd. Aren't proofs *paradigms* of arguments on the hoof? When the workaday theoretician labours to prove a new theorem of logic or of mathematics is she not marshalling arguments on the hoof? So it would seem that as a general model of deductive arguments on the hoof, **PC**-syllogisms are significantly crimped.

Whether the same fate awaits Aristotle himself depends on whether we think that, undefined or not, Aristotle's validity was "classical", i.e., as standardly defined in orthodox systems such as **PC**. Scholars are divided on this question. It is a question that takes us far from our purposes here. We shall let it be, except to say in passing that even if Aristotle's validity is standard, the ancient syllogistic may fairly reclaim the high ground over **PC**-syllogisity as an account of arguments on the hoof.[12]

Earlier we met with another complaint against standard systems of logic. Standard systems, it was said, have as their target concept such stripped-down things as logical truths or the implication relation. And it was asked, rhetorically, what headway the theorist of argument or of inference could make with a conceptual bag of tricks restricted to the logical truths or the implication relation or others derivable from them. The answer, as we now see, is one that pulls the question's rhetorical fangs. For in fact in some standard systems, such as **PC**, the concepts of logical truth, implication *and* argument validity are interdefinable: A logical truth in **PC** is a tautology and nothing else; Φ implies Ψ in **PC** if and only if $\ulcorner$If Φ then $\Psi\urcorner$ is a tautology; and an argument is valid in **PC** if and only if its corresponding conditional is a tautology.

[12]cf. Normore [1993, 447]. For the view that even if Aristotle's validity is standard the ancient syllogism may reclaim the high ground, see Woods [2001].

Comparisons with Aristotle aside, **PC** aspires to be a method for revealing the *logical forms* of a large class of arguments on the hoof. They are the arguments whose propositional constituents can be given a representation in **PC**. Derivatively, they are arguments which can be represented by **PC**-sequences of **PC**-sentences. We have already seen how this works for simple quantifier-free sentences of English. They are represented by atomic sentences of **PC**. It is important to say what their representation amounts to. There are two ways of seeing the process. One involves assuming that the vocabulary, sentences and sequences of **PC** constitute an *artificial language* into which the vocabulary, sentences and sequences of a natural language are translated. In the other way of thinking of it, representation is an operation directly on a natural language itself (or fragment of it). By these lights, representation is an abstraction from language on the hoof. Seen the first way, the English sentence, "Mary is angry" is translated as the atomic sentence p_1 (say) of **PC**. Seen the second way "Mary is angry" is *reconstructed* as p_1 (say), where all of the interior grammatical structure of "Mary is angry" is suppressed. Simple sentences are a special case. When we suppress all of their internal structure, they leave no lexical trace. Labels are required, one each for each suppressed sentence, and this labeling function is performed by the atomic sentence-letters $p_1, p_2, \ldots, p_n \ldots$. Thus p_1 denotes some or other simple sentence whose grammatical interior has, for theoretical purposes, vanished. Similarly for p_2, p_3, and so on.

With molecular sentences, reconstruction becomes a more relaxed way of talking. If we take an example the English sentence, "If Mary is angry then Bill is sad", it is reconstructed as "If p_1 then p_2" if p_1 reconstructs "Mary is angry" and p_2 "Bill is sad." We thus see part of the logical form that "If Mary is angry then Bill is sad" actually has. Reconstruction, then, is the mapping of a sentence on the hoof into its logical form.

There is not much to be said about which is the better way of taking representation, either as reconstruction or translation. Either way, what we get is a *formalization* of something in English (say) in the language of **PC**.

6 FORMALIZING NATURAL LANGUAGE

We are now in a position to determine whether an earlier promissory note can be redeemed. In chapter one of this volume it was proposed to take logic as formal idealized description of the behaviour of a logical agent. One of the reasons for so proposing is the desire to have a conception of logic that engages practical reasoning in a fruitful way. Since practical reasoning can be taken as reasoning done by a practical agent, and since practical agency has been described as an agent's implementation of scarce resource strategies in a cognitive economy, *agency-type* lies at the theoretical heart of a logic of practical reasoning.

We remarked at the beginning of the present chapter that the practical-agent conception of logic seems in its face to be hostile to the dominant notion of logic since Frege and Peirce. In all its sub-domains — proof theory, model theory, recursion theory and set theory — standard logic is indifferent to agency, except in the most contingent of ways. Standard deductive logic is a disciplined description of target properties of linguistic structures, of structures in which pragmatic factors of agency, time, context and the like are suppressed. It is time that some attempt has been made by modern logicians to adopt standard logics with a view to making their rules useful for real-life agents. Deontic logic is a case in point. In its choice of model operators for obligation and permission and their contraries, there is a clearer adumbration of agency to be seen. But it remains the case that deontic logic is a formal description of the properties of propositional structures.

Of course, it can only be expected that part of what a logical agent will do is ascribe or be guided by the logical properties of propositional structures. His behaviour as a logical agent won't likely be indifferent to considerations of implication, consistency, incompatibility and the like. To that extent the agency-conception of logic can be said to subsume the logics of propositional structures, thus diminishing the apparent hostility between them. The rivalry would be further reduced if it could be shown that the logic of propositional structures elucidates argument on the hoof. Suppose it were the case that standard formal logic — **PC**, say — afforded a principled way of determining the presence of its own target properties in the product-arguments of process-arguments on the hoof? That would be more mitigation still, of the rivalry between the agency conception and the propositional-structures conception of logic. In the present section we explore this possibility.

Not even the most ardent or doctrinaire mathematical logician believes that human reasoners should or could abandon their mother tongues. Frege and Peirce would grant that for everyday concerns one's mother tongue is the unavoidable medium in which they are to be addressed. Frege and Peirce would also have insisted is that human reasoning in a human language is something that lies beyond the ambit of logic.

This is one of the things that disappoints (actually, infuriates) informal logicians about mathematical logicans. It is their insistence that "ordinary reasoning" has no logic. However some formal logicians are heedful of such disappointment and are minded to respond to it in the following (and somewhat conciliatory) way.

The basic idea is this. Take an episode of human reasoning or argument-making which has been transacted in a piece of (say) English. Subject to certain constraints, this chunk of English can be tightly paired with a counterpart bit of the language of **PC**. The tightness of the fit is consequential. It allows us to say that if the **PC**-structure has one of our target

logical properties, so too does the English structure have it. The process of finding an English structure's counterpart in **PC** is its *formalization*. When a formalization works as it should, it maps an English structure to its *logical form* in **PC**. And, again when the formalization has been properly contrived, the property possessed, by an English structure's logical form is *reflected backwards* onto the English structure itself.

To better see how this works, here are the standard formalization rules for English and **PC**.

1. Simple bivalent sentences of English, and only they, are formalized by arbitrarily selected atomic sentences of **PC**.

2. The English construction 'not' is formalized by '$\neg$'.

3. The English constructions 'and', 'or', 'if ... then' and 'if and only if' are formalized respectively by '$\wedge$', '$\vee$', '$\supset$' and '$\equiv$'.

In this way, the formal logician thinks that he has the means to reassure his critic. In its most basic sense, the assurance is this: If we can formalize some reasoning (or arguing) in English in some or other logical system, then, since formalization has the backwards reflection property with respect to target concepts, if the logical form instantiates that concept, so too does the English construction.

Essential to the case for the backwards reflection of target properties is that the connectives of English stand in the following truth conditional relationship. Let C be an English connective which is a candidate for formalization; and let K be its formalization in **PC** if it has one. Then C is properly formalized as K, provided that for any K-sentence that is false in **PC**, its corresponding C-sentence in English is also false. In other words, the falsity of a K-sentence is sufficient for the falsity of the corresponding C-sentence. But not conversly; the falsity of the C-sentence is not sufficient for the falsity of the K-sentence; nor is it the case that the truth of the respective C-sentence and K-sentence; suffices for the truth of the other. So K can be the absolutely right formalization of C in **PC** without there being any need for C and K to be logically equivalent to one another, to say nothing of synonymous. We see, then, that the fact that our **PC** connectives do not mean the same as their English counterparts or do not capture at least one of their meanings (in case they happen to be ambiguous) has nothing whatever to do with whether the formalization rules have the backwards reflection property with regard to our target concepts.

But what is to be done when an English connective C fails this test? Then the rule is that the connective in question does not *have* a formalization in **PC**, nor does any sentence of English in which C occurs. The rule can now be explicitly stated.

The Connective Rule — Where C is an English connective and K its counterpart in **PC**, K is the formalization of C in **PC** if the falsity of any K-sentence implies the falsity of its corresponding C-sentence; otherwise C has no formalization in **PC**.

Backwards reflection is really quite remarkable. It shows that formalizing a content*ful* argument in English as a content*less* argument in **PC** allows us to determine with certainty whether the *English* argument is valid.

Some people are of the view that it is too good to be true. To see what their reservation comes down to, it is necessary to emphasize that there are constraints on what can be inputs to our formalization rules. For example, we are not allowed to apply these rules to interrogative sentences of English, nor are we allowed to formalize molecular sentences of English as atomic sentences of **PC**. And we are not permitted to formalize any connective of English other than 'not', 'or', 'and', 'if ... then', and 'if and only if', and those that can be defined in terms of these connectives.

Consider, in particular, the rule that only simple sentences of English can be mapped to atomic sentences of **PC**. Are there any other constraints on this atomic rule? Consider the argument:

1. If Sarah has been awarded the first university degree, then Sarah is a bachelor.

2. If Sarah is a bachelor then Sarah is an unmarried man.

3. Therefore, If Sarah has been awarded the first university degree, then she is an unmarried man.

If we now apply the formalization rules to this argument, we see that its logical form in **PC** is

1*. If p then q

2*. If q then r

3*. Therefore if p then r.

This matters. Here is an English argument with a valid logical form in **PC**. But the English argument is invalid. We wanted validity to have the backwards reflection property, but the present example shows that it doesn't.

We can solve this problem by noticing that our English argument equivocates on the ambiguous term "bachelor". In premiss (1) it means one thing, and in premiss (2), it means something quite different. This suggests a way out of our difficulty. We can impose upon the formalization of English sentences the *Disambiguation Rule.* The rule says that for any expression of English which has more than one meaning, its different meanings require a

mapping for each sentence in which it occurs to different expressions of **PC**. Applying this rule to our present example, we see that the correct logical form of our English argument is

a) If p then q

b) If s then r

c) Therefore, if p then r.

This gives us the desired result. The logical form is invalid. So it would appear that when we add the Disambiguation Rule to our rules of formalization, validity does indeed have the backwards reflection property; for we no longer have an argument which is valid in **PC** but invalid in English.

What about *in*validity? Does it too have the backwards reflection property? Consider the following argument.

1. The shirt is red.

2. Therefore, the shirt is coloured.

The premiss is a simple sentence of English which entails the conclusion, also a simple sentence of English. Its logical form in **PC** is

1*. p

2*. Therefore, q

which is invalid.

Here is a second case to consider.

a) The figure is a triangle

b) The figure is a circle

This is an inconsistent set of sentences in English. But the logical form of this set in **PC** is $\{p, q\}$, and $\{p, q\}$ is a consistent set.

In the first case, invalidity fails to have the backwards reflection property. In the second case, consistency fails to have the backwards reflection property. But we want *all* our target properties to satisfy the backwards reflection condition. We want this because we want **PC** to be useful in the appraisal of real-life reasoning and real-life argument.

As it happens, we can recover the backwards reflection property with regard to invalidity and consistency if we agree to impose a further condition on our formalization rules. This is the

Logical Inertia Rule
Simple sentences of English to which the formalization rules apply may not either imply one another or be inconsistent with one another. In other

words, the simple sentences that are inputs to the formalization mechanism of **PC** must be *logically inert*.

It seems, then, that we have recovered by backwards reflection of validity by imposing the Disambiguation Rule, and likewise that we have recovered the backwards reflection of invalidity and consistency by imposing the Logical Inertia Rule. Even so, there is a cost to these recoveries.

Let us deal first with the cost of imposing the Logical Inertia Rule. This requires us to be able to recognize implications and inconsistencies between simple English sentences in a principled way. This is equivalent to saying that we must have a *theory* of implication and inconsistency for English. But this is what our formalization rules were supposed to provide. **PC** would analyze the properties of implication and consistency, and our formalization rules would reflect them back into English. So **PC** together with the formalization rules would be a theory of implication and consistency for English. But, as we now see, we can't run the formalization rules until we have a theory of implication and consistency for English. And we don't have a theory of implication and consistency for English until we've executed the formalization rules. So we have a *bootstrapping problem*.

Faced with this kind of difficulty, most logicians have in effect withdrawn the Logical Inertia Rule and, in so doing, have abandoned the hope that invalidity and consistency would have the backwards reflection property.

The Disambiguation Rule is also a serious matter. If it goes, then we lose the backwards reflection of *validity*. So the question now is whether we are able to apply this rule in a principled way. The answer is no. We do not yet have a theory that permits us to recognize ambiguity in the general case. (In fact, a good many philosophical problems have turned out to have resulted from *undetected* ambiguities.) This leaves us with Hobson's Choice. Either we can give up on the backwards reflection of validity or we can try to apply the Disambiguation Rule in other than a principled way, that is to say, *intuitively*. Most logicians opt, in effect, for the second option. This also matters. No one doubts that a native speaker of English is adept at recognizing large numbers of violations of the Logical Inertia Rule. This is part of what fluency in a language consists in — the intuitive ability to notice elementary logical connections. The same can be said for the fluent speaker's untutored capacity for making and recognizing logical deductions in his own language. The question is whether we have theories of such skills, and if so what they would look like. Gerald Massey has suggested that one place to look for theories of these capacities is in the Natural Logic of writers such as Lakoff (Massey [1975], [1975a], [1981]). I do not know whether Massey has retained his enthusiasm for Natural Logic over the years; but it is now clear, if it wasn't before, that Natural Logic has not attracted anything like a large and settled theoretical consensus. Natural Logic aside, two other points should be made. One is that nowhere in the capacious

writings of informal logicians do we find any attempt to construct theories of (logical) implication and (logical) consistency for natural languages. The other is that monitoring the Disambiguation Rule and the Logical Inertia Rule is not the business of (or within the capacity of) formal logic. If formal logic is to offer any principled guidance to natural language reasoners and arguers, it will be able to do this only on the basis of an irreducibly informal deployment of the requisite constraints on formalization. Informality is prior to formality, and not displaceable by it.

In their collective (though largely tacit) decision to abandon hope of backwards reflection except for logical truth (or taulolgousness) and argument-validity, standard logic has given up significantly reduced its usefulness in the appraisal of real-life argumentation. The gap between the agency conception and the propositional-structures conception of logic widens accordingly. Still a test for validity is nothing to sneeze at (assuming that we have any ambiguities under control).

We note in passing, subject to this same point about ambiguity, that the **PC** validity-test does *not* require that the formalization of English be all that tight. Tightness was required for the backwards reflection of properties, such as invalidity, that logicians have already given up on. In particular, we need not require that the English constructions that get mapped to atomic **PC**-sentences be simple (that is, connective-and quantifier-free). Consider a case. Let A be a **PC**-argument in the form of *modus ponens*; and let $e_1, \ldots,$ e_n be an arbitrarily selected set of English sentences, as richly structured as you like. Finally, let our formalization procedures take some of the e_i to the atoms of A. A is valid when for no interpretation is the conclusion false and the premisses conjoined true. Since the e_i have truth-values, and since the validity of A is indifferent to the truth values of its atoms, then in mapping e_i to the atoms of A, all that matters about the e_i is their truth values; and yet it is precisely these that A is indifferent to if it is valid. Thus, subject to a proviso, *any* truth-valued sentence of English can be mapped to any atom of **PC** in a test for validity (or logical truth), provided that the Connective Rule is honoured.

Put in Massey's way, the Asymmetry Thesis is that we do not have a theory of invalidity for natural language arguments. (Massey [1975] and [1975a]) Put in the language of formalization rules, the thesis is that judgements of validity do not require tight atomic formalizations, whereas judgements of nonvalidity require atomic formalizations to be tight beyond our present theoretical capacity to provide.

7 OTHER ASPECTS

If **PC** is motivated by an interest in capturing for ordinary arguments the property of validity, it also enjoys other motivations. One is broadly *proof-*

theoretic, and it underwrites an interest in reducing *mathematical reasoning* to exactly formulated mechanical rules. In this it was supposed that all underlying assumptions would be expressly declared and all proofs would be rigorous. The proof-theoretical approach was developed by Frege (1862–1925), Peano (1858–1932), Hilbert (1848–1925), Russell (1872–1970), Herbrand (1908–1931), and Gentzen (1909–1945). In Hilbert's eyes, the logic would subject mathematical reasoning itself to the scrutiny of mathematical theory. In this way, Hilbert sought to give a mathematical justification of a (moderately) infinitistic approach to mathematics. Much of the work produced under this motivation has come to be known as *formal proof calculi*, and of these, Frege's *Begriffsschrift* [1879], was the first to produce a satisfactory notation and theory of proof for what is known as *first order logic*, which is **PC** extended by the introduction of predicate terms, quantifiers, and individual terms accessible to quantification.

A third basic motivation for the new logic may be called *model-theoretic*. It is a highly abstract motivation, in the same spirit as that which underlies the upper reaches of set-theoretic mathematics. It is essentially an interest in abstract structures and in the conditions which such structures must satisfy in order to stay interesting. Leading this model-theoretic effort, were Schröder (1841–1902), Löwenheim (1878–1957), Langford (1895–1964), Gödel (1906–1978) and Tarski (1901–1985).

If a reader's interests run to deductive argument and deductive inference as such, it is not surprising that he would arrive at a feeling of alienation running deep in the marrow of proof theory and model theory. If the disappointed argumentation theorist is a sober-minded fellow, he will quickly see that much of the most interesting and important work done in these branches of logic will have little direct relevance for him. If he is a bit of a pepperpot, he is likely to be cranky about the whole enterprise, ready to damn all logic as an imperious falsifier of the delicate tissue of argumentative and inferential practice. It is not our purpose in this chapter either to develop or adjudicate such complaints. All the same, a certain openness may seem appropriate, in as much as proof theory deals with *proof* and in as much as mathematical reasoning is, after all, reasoning.

Both the proof theoretic and the model theoretic approaches are approaches that can be taken to **PC**. In fact, the valuations approach to **PC** taken in these pages is (a very modest) *part* of a model theoretic methodology. But proof theory and model theory came into full flower in the richer systems known as the *predicate calculus*, **PredCal**, to an elementary exposition of which we now turn.

It is apparent that the three approaches converge in **PC** in a rather fundamental way. Each has a stake in specifying truth conditions on the schema: $\ulcorner\Phi_1, \ldots, \Phi_n\urcorner$ *logically implies* Ψ. From the point of view of argumentation theory, the condition would be that every argument of the form $\langle\{\Phi_1, \ldots \Phi_n\}, \Psi\rangle$ be valid. A proof theorist would require that there is a

proof of Ψ from $\{\Phi_1, \ldots, \Phi_n\}$ in a *standard proof calculus*. A fancier of model theory would require it to be true that whenever the Φ_i are all true *in a structure*, Ψ is also true there.

8 PREDICATE LOGIC

In the discussion of **PC**, we noted an attractive concurrence between logical implication, valid argument and proof. As we have seen, arguments in English can be said to be valid should it be the case that it embeds a valid **PC**-form. **PC**-forms are extremely simple structures, however. This matters in two related ways. Having the requisite such structure delivers the goods (ambiguity problems aside) only for certain of the logical properties people are interested in. Argument-validity is one such, true enough; but even so, there are large classes of valid arguments in English whose validity cannot be captured by truth functionally valid forms. A case in point is our previous example,

1. The shirt is red.
2. Therefore, the shirt is coloured.

Another is,

a. Harry is male.

b. Therefore, someone is male.

These examples turn on interconnectives of syntactic elements that **PC** has no capacity to recognize. In case one, the deduction turns on the pair of *predicates* {"red", "coloured"}, and in the second case it turns on the pair {"Harry", "someone"} of a *name* and a *quantifier*. These facts make it extremely tempting to consider whether a grammatically richer language would make for correspondingly richer logical forms with which more realistically to capture the logic of real-life arguments. With this possibility in mind, we now turn to the predicate calculus.

Like **PC, PredCal** is a sequence $\langle \mathbf{G}, \mathbf{I^*} \rangle$ of a grammar and a set of interpretations or models. Here, too, **G** comes in various degrees of complexity.

We begin with the vocabulary of **PredCal**. Its vocabulary contains the vocabulary of **PC**, to which are added (1) countably many *individual constants*, a_1, a_2, ... and (2) countably many *individual variables* $x_1, x_2, \ldots$. The individual constants function as proper names of individual objects, and the individual variables as pronomial expressions for those objects. Such constants and variables constitute the *terms* of **PredCal**. The vocabulary also includes countably many predicate constants, $F_1^1, F_2^1, \ldots, F_1^n$,

$F_2^n, \ldots, F_1^{n+1}, F_2^{n+1}, \ldots$. The superscript gives the polyadicity of the predicate, that is, it marks the number of *places* it has; and the subscript serves to individuate all predicates of given degree of polyadicity, for each degree of polyadicity. As with the atomic sentences of **PC**, the individual and predicate constants of **PredCal** are also atomic.

So far, we have extended as follows the vocabulary of **PC** to arrive at the vocabulary of the new system **PredCal**.

I. The atomic sentences p, q, r, s, t, $p_1, \ldots, p_n, \ldots$

Ia. The individual constants or names $a_1, a_2, \ldots, a_n \ldots$

Ib. The predicate and constants $F_1^1, \ldots, F_n^1, \ldots, F_1^n, \ldots, F_n^n, \ldots$

Ic. The individual variables $x_1, x_2, \ldots, x_n, \ldots$

IIa. The connectives $\neg$, $\wedge$, $\vee$, $\supset$, $\equiv$

IIb. The punctuators (and), and sequence markers $\langle$ and $\rangle$

IIc. The quantifier $\forall$.

We now define a *singular term* of **PredCal**: Any individual constant or variable is a singular term of **PredCal**.

The *sentences* of **PredCal** (or the **PredCal** formulas) are inductively characterized as follows. The *prime* **PredCal** *formulas* are the atoms of **PC**; so is any expression $\ulcorner\Phi(\alpha_1, \ldots, \alpha_n)\urcorner$ where Φ is a n-ary **PredCal** predicate constant and $\alpha_1, \ldots, \alpha_n$ are singular terms. The **PredCal** formulas are the least class including the prime formulas that meets the conditions:

1. If Φ is a **PredCal** formula, and α is a **PredCal** variable then $\ulcorner\forall\, \alpha\, (\Phi)\urcorner$ is a **PredCal** formula.
2. If Φ is a **PredCal** formula so is $\ulcorner\neg\Phi\urcorner$.
3. If Φ and Ψ are **PredCal** formulas, so are $\ulcorner\Phi \wedge \Psi\urcorner$, $\ulcorner\Phi \vee \Psi\urcorner$, $\ulcorner\Phi \supset \Psi\urcorner$ and $\ulcorner\Phi \equiv \Psi\urcorner$.

In a formula $\ulcorner\forall\alpha(\Phi)\urcorner$, α is said to be the variable of the quantifier.

As was the case with **PC**, we do not carry on our discussion of the language of **PredCal** in the language of **PredCal**. Our metalanguage is English augmented by some special metalinguistic symbols. In particular, we are now employing 'α' and 'β' as metalinguistic variables ranging over singular terms and individual variables of the object language, **PredCal**. We also make the convention that Φ, Ψ, χ, with or without accents are variables ranging over **PredCal** predicates as well as **PredCal** formulas. We now define:

The scope of a quantifier. If $\ulcorner\forall\alpha(\Phi)\urcorner$ is a formula then the occurrence of the formula Φ in $\ulcorner\forall\alpha(\Phi)\urcorner$ is the scope of $\forall$ in $\ulcorner\forall\alpha(\Phi)\urcorner$.

Freedom and bondage. An *occurrence* d of variable α is *bound* in a formula Φ iff either (1) d follows an occurrence of a quantifier in Φ; or (2) d occurs in the scope of such a quantifier, otherwise d is a *free*, occurrence of α in Φ.

A *variable* α is *free* or *bound* in a formula according as Φ contains free or bound occurrences of α. A variable can be both bound and free in a formula.

(c) *Freedom of a singular term for a variable.* We define the conditions under which Ψ is a *substitution instance* of Φ: If Φ is a formula, then Ψ is a substitution instance of Φ just in case Ψ is the formula $\ulcorner S^{\beta}_{\alpha}(\Phi)\urcorner$ that results from substituting the singular term α for all free occurrences of a variable β in Φ. It is not always the case where Φ, α, β exist that $\ulcorner S^{\beta}_{\alpha}(\Phi)\urcorner$ is a substitution instance. But, as our exposition here is elementary, we will overlook this fact, and we will suppose restrictions on Φ, α, β such that $\ulcorner S^{\beta}_{\alpha}(\Phi)\urcorner$ is always a formula. (This is nicely dealt with in Bell and Machover [1977, chapter 2 section 3].)

Let α be a singular term and β a variable of a formula Φ. Then α is *free for* β in Φ according as the following conditions are met: either

(i) α is itself a variable and no free occurrence of β in Φ lies within the scope of a quantifier such that α is the variable of that quantifier; or

(ii) α is a name.

Closedness. Φ is a *closed* formula if and only if Φ is a formula no variable of which occurs free. Any formula that is not closed is *open*. Closed formulas are *sentences*.

Quantifier closure. Let Φ be any formula containing only, at most variables $\alpha_1, \ldots, \alpha_n$ free; then the formula $\ulcorner\forall\alpha_1 \ldots \forall_n(\Phi)\urcorner$ is a *quantifier closure* of Φ.

Finally, to complete our description of the language of **PredCal** we introduce the expression '$\exists$'. We shall say that $\exists$ is definable from $\forall$ in the special sense that any formula

$$\exists\alpha(\Phi)$$

is strictly equivalent to

$$\neg\forall\alpha\neg(\Phi),$$

where Φ is a formula and α a variable. As will be recalled, $\forall$ is suggestive of the English "all" and therefore is called the *universal* quantifier; $\exists$, suggestive of the English "some" and "at least one," is called the *existential* quantifier.

Formalizing sentences of English. Although these considerations play no formal role in the grammar of **PredCal**, it should be evident that one of the goals of our proceedings is, as we have suggested to represent in **PredCal** more of English than we could represent in PC. Thus we might expect that under a suitable formalization policy a simple subject-predicate sentence of English, say,

(1) Paris is charming.

will be represented by a **PredCal** prime formula, say,

(1*) $F_1^1(a_3)$

where, "F_1^1" represents the one place predicate "is charming" and "a_3" represents the name "Paris". And an ordinary sort of general sentence of English, say,

(2) For any object whatever either it is not a man or it is a man,

could be associated with the closed formula

(2*) $\forall_1(\neg(F_4^1(x_1)) \vee F_4^1(x_1))$.

(2*) can be read "For any object, call it x_1, either it is not the case that x_1 is F_4^1, or x_1 is F_4^1".

In the same way, one could hope that, just as (2) is a necessary truth of English, so too would (2*), or whatever other **PredCal** formula might represent it, be a **PredCal** logical truth.

Model theory. In its most basic form, model theory undertakes to do two related things. It tries to elucidate the relationship between language and the world (or between language and our experience of the world), by construing truth as a relation between a sentence and a model. A *model* is a structure composed of a *domain* (of individuals) together with an *interpretation*, where this latter is an assignment of the appropriate values to the basic constituents of the grammar. The assignment associates individuals with individual constants of the language in question, functions with functors or function symbols, and properties and relations with predicate symbols. A prime formula free of variables is true in the structure (equivalently, the structure is a *model* of the sentence.) if and only if the individuals assigned to the sentence's names or individual constants bear the relation assigned to the sentence's predicate symbol. Model-theoretic structures such as these assume that experience arises from an independently existing reality, of which they are a kind of mathematical or set-theoretic idealization.

The second task of model theory is to specify two further relations between language and the world so as to make possible the elucidation of the

concept of logical consequence or entailment. The relation of logical consequence from a set of sentences $\sum$ to a sentence Φ is said to obtain just in case every structure or model in which all the sentences of $\sum$ are also true Φ is one in which Φ is also true. This is meant to capture the intuitive idea that a proposition B follows from propositions $A_1, \ldots, A_n$ if and only if it is in no sense possible that the A_i are true and yet B is not true. We now turn to some details of model theory.

We characterize a **PredCal** *structure* $\sum = \langle D, I_o \rangle$ such that $D \neq \varnothing$ is the *domain* of $\sum$ and I_o an interpretation function. By the interpretation function I_o we have it that

(1) every **PC**-atom of **PredCal** is presumed to denote either T or F, but not both;

(2) every name is presumed to denote exactly one member of D;

(3) every unary or one-place predicate is presumed to denote the sets of elements from D;

(4) every n-ary predicate ($n \geq 2$) is presumed to denote a set S of ordered n-tuples of objects taken from D (that is, of n-membered sequence of objects taken from the D). These sets of n-tuples are called n-place relations;

We also remark that the connectives of **PredCal** behave just as they did in **PC**; and that the formulas of **PredCal** will either be *satisfied* (or not) by *countably infinite* sequences of objects from D. We now turn to the notion of a formula's being satisfied by a countably infinite sequence out of the domain of an interpretation.

The individual variables each bear a unique index. Here is why. Let Φ be any **PredCal** formula in which α is a variable occurring free. Consider any of the infinitely long sequences of members of D — call it σ. Since σ is a sequence, its elements occur there in a given order (they might, of course, occur in quite a different order in another such sequence). So it is meaningful to speak of the kth member of σ, where k is any positive integer. As for Φ, we shall say that its variable $\alpha = x_i$ denotes the i^{th} member of *every* sequence in S. It is very important to note that we did *not* say, just now, that it denotes the ith member of σ and *that same member* in every other sequence in S. Rather we said that, for all sequences in S, its ith member, whichever elements they severally are, is denoted by α, free in Φ.

For the purposes of illustration let us assume D to be the (unordered) set of all the positive integers, augmented by the three charming towns of Banff, Picture Butte and Standoff. Consider any infinitely long sequence σ of elements from D. We define a function, d^{σ}, from singular terms into the domain D: if α is a name, then $d^{\sigma}_{k(\alpha)}$ is the element of D that, the name α names. If α is a variable and k is the index of the variable, then d^{σ}_{α} is any

element x in D such that x is the kth element of σ, where σ is any sequence in S.

It is worth emphasizing that variables need not identify what they denote; rather they *identify* only the sequence-place of what they denote. And they *denote*, for each and every sequence, the object that there occurs in some given fixed place.

We are now in a position to define the concept of *satisfaction of a formula by an infinite sequence of D-elements.* Let $\Sigma = \langle D, I \rangle$ be a model. Such sequence, σ, satisfies Φ in Σ actually: relative to a structure or model $\Sigma = \langle D, I \rangle$ since D has a role to play as well (or Σ-satisfies Φ) according to the following conditions being met.

(1) If Φ is a **PC**-atom, then Φ Σ-satisfies σ iff **I** provides that Φ denotes the truth value T.

(2) If Φ is a prime formula of **PredCal**, $\ulcorner \Psi(\alpha) \urcorner$, where Ψ is a unary predicate and α is a singular term then σ satisfies Φ iff d^{σ}_{α} is a member of the set which (under Σ) is denoted by Ψ.

(3) If Φ is a prime formula of **PredCal**, $\ulcorner \Psi(\alpha_1, \ldots, \alpha_n) \urcorner$, where Ψ is an n-ary predicate ($n \geq 2$) and $\alpha_1, \ldots, \alpha_n$ are singular terms, then σ satisfies Φ iff $\langle d^{\sigma}(\alpha_1), \ldots, d^{\sigma}(\alpha_n) \rangle$ is a sequence that is a member of the set of n-tuples which (under Σ) is denoted by the n-ary predicate Ψ.

(4) If Φ is the **PredCal** formula $\ulcorner \neg \Psi \urcorner$ then σ satisfies Φ iff σ does not satisfy Ψ.

(5) If Φ is a **PredCal** formula $\ulcorner (\Psi \vee X) \urcorner$ then σ satisfies Φ iff either σ satisfies Ψ or σ satisfies X or both. If Φ is $\ulcorner (\Psi \wedge \chi) \urcorner$, then σ satisfies Φ iff σ satisfies Ψ and σ satisfies χ. If Φ is $\ulcorner (\Psi \supset \chi) \urcorner$ then σ satisfies Φ except when σ satisfies Ψ and does not satisfy χ. If Φ is $\ulcorner (\Psi \equiv \chi) \urcorner$, then σ satisfies Φ iff either σ satisfies both Ψ and χ or σ satisfies neither.

(6) If Φ is a **PredCal** formula $\ulcorner \forall \alpha (\Psi) \urcorner$, where α is the variable whose index is the integer k, then σ satisfies Φ iff every countably infinite sequence (of elements of D) that differs form σ at most only in its kth element is a sequence that satisfies Ψ.

Readers may have noticed that, by the first two clauses of the account of satisfaction, it is intuitively correct to say that true **PC**-atoms and true **PredCal** primes are satisfied by every sequence, even by those sequences no member of which is an image under the function, d^{σ}, of any constituent parts of the **PC**-atom or the prime formula of **PredCal**. But this is as it should be. We want no true formula to be countersatisfied by any sequence

of S, and any formula that is not countersatisfied by such a sequence is satisfied by it.

It may not be entirely clear how clause (5) achieves the desired result, viz., that a universally quantified sentence holds for any model exactly if the scope of the quantified sentence is true of every object in the domain of the model. To see how this is accomplished, let us recall our homely geopolitical-arithmetical domain of a few pages back. It is the unordered *set*

$$S = \{\text{Banff, Picture Butte, Standoff, 1, 2, \ldots, n, \ldots}\}.$$

Assume now a structure of **PredCal** which seeks to preserve the truth that, relative to this domain, everything is either a town or an integer, or, as we shall say for short, that everything is a *tinteger* S is the set of all tintegers. Let I be such that $I(F_1^1) = S$. Then there will be sentence of the form

(1) $\forall x_k(F_1^1(x_k))$,

that we shall want to be Σ-satisfied.

Clause (5) says that this sentence is Σ-satisfied by an infinite sequence σ provided that its scope, i.e. the formula

(2) $F_1^1(x_k)$,

is satisfied by every infinite sequence differing from σ (if at all) only in its kth member. Put another way, σ satisfies (1) iff every sequence σ satisfies (2).

Suppose now that σ is the *sequence*

$$\langle \text{Banff, Picture Butte, Standoff, 1, 2, \ldots, } n, \ldots \rangle$$

and let us ask whether σ Σ-satisfies the formula

(3) $\forall x_3(F_1^1(x_3))$.

We can see at once that σ differs *from itself* in at most its k^{th} place (i.e. its third place). In fact, itdiffers there not at all. So the first thing to check is whether σ *itself* satisfies the scope of (3). Does σ satisfy

(4) $F_1^1(x_3)$?

Our interpretation assigns "x_3" to the third element of σ, that is, to Standoff $d^3(x_3) =$ Standoff. Hence $< d^3(x_3)$ is a member of the set S, that is $I(F_1^1)$, because it is a town, hence also a tinteger (i.e. a town or an integer). So σ satisfies "$F_1^1(x_3)$".

Next we consider any other sequence of elements of D differing from σ in exactly its third place. Any such sequence is just like σ except that it will have as its third member either Picture Butte or Banff or the integer 1, or

2 or ..., or n, or .... It is clear in particular that in each such sequence the third element will always be either a town or an integer; hence every such sequence will differ from σ only by having at its third place a tinteger different from the tinteger that occupies σ's third place. This is the only sort of difference that can occur, and it is an inessential one from the point of view of the interpretation of "$F_1^1(x_3)$". Every such sequence satisfies "$F_1^1(x_3)$" and, so, by clause (5), σ satisfies "$\forall x_3$ ("$F_1^1(x_3)$)". It should now be clear that σ could not have Σ-satisfied this formula were it not precisely for the fact that every element of our D is an tinteger — a wholly desirable and intuitive state of affairs after all.

We must also remark that Σ does not actually require the use of infinite sequences of D-elements. In some treatments of so elementary a grammar as that of **PredCal**, a satisfaction relation can be defined without the use of infinite sequences (Shoenfield [1967]). Thus our semantics are richer than they need to be for present purposes.

Logical Truth in the Model Theory of PredCal. Now that sequence-satisfaction has been sorted out, the other semantic properties of **PredCal** come fairly easily

We may say that Φ is a *satisfiable* **PredCal**-formula just in case there is at least one model or structure on which at least one countably infinite sequence of elements of the domain satisfies Φ.

A set Γ of **PredCal** formulas is *simultaneously satisfiable* just in case there is at least one structure, and at least one countably infinite sequence from its domain, which satisfies every **PredCal** formula in Γ.

A **PredCal** formula Φ is *true in a model* $\Sigma = \langle D, I \rangle$ just in case every countably infinite sequence from its domain satisfies Φ (note, then, that not every sequence that satisfies a formula makes it true).

Φ is *false in* a model $\Sigma = \langle D, I \rangle$
just in case no countably infinite sequence from D satisfies Φ.

A prime formula may be neither true in Σ nor false in Σ, for some Σ. This is so, when Φ is a prime formula some of whose variables occur free. However, we note that in a given structure every closed formula must be ? by all or by none of the sequences, and hence must be either true or false. For example it may be that a variable occurs free, yet the formula is, say, true in Σ. For instance, $\ulcorner P_x \urcorner$ where I assumes D to P ($I(P) = D$) would be true in Σ.

Any model or structure Σ in which a formula Φ is true is said to be a *model for* Φ.

Any model or structure Σ in which a formula Φ is false is said to be a *countermodel* for Φ.

Any assignment Σ is a model (countermodel) for a set Γ of **PredCal** formulas just in case every formula in Γ is true in (false in) Σ.

A formula of **PredCal** is a logical truth (or a valid *sentence*) just in case every model or structure is a model for Φ.

A **PredCal** formula Φ (or set thereof) *semantically entails* a **PredCal** formula (or set thereof) just in case there is no model or structure that is a model for the former that is not a model for the latter. (Thus, in particular, the empty set of formulas semantically entails any logical truth or valid formula).

Proof Theory. An axiomatization *A***PredCal** of **PredCal** is now sketched. The language and Formation Rules are as laid out previously. The definitions of *scope*, *freedom*, *bondage*, *freedom for*, *closedness* and *quantifier closures* are carried over intact.

The *axioms* of *A***PredCal** are specified by the following list.

($A1$): $\ulcorner(\Phi \supset(\Psi \supset \chi))\supset((\Phi \supset \Psi)\supset(\Phi \supset \chi))\urcorner$

($A2$): $\ulcorner\Phi \supset(\Psi \rightarrow \Phi)\urcorner$

($A3$): $\ulcorner(\neg\Phi \supset \neg\Psi) \supset (\Psi \supset \Phi)\urcorner$

($A4$): $\ulcorner\forall\beta\Phi \supset S^{\beta}_{\alpha}(\Phi)\urcorner$, provided α is free for β in Φ

($A5$): $\ulcorner\forall\alpha(\Phi \supset \Psi)\supset(\Phi \supset \forall\alpha\Psi)\urcorner$ provided α is not free in Φ

There are two *Transformation Rules.*

($T1$): *Detachment.* For any **PredCal** formulas Φ and Ψ, Ψ is deducible from the set $\{\Phi, \ulcorner\Phi \subset \Psi\urcorner\}$

($T2$): *Generalization.* For any **PredCal** formula Φ and **PredCal** variable α, $\forall\alpha(\Phi)$ is deducible from Φ.[13]

As before, it is easy to characterize an *A***PredCal**-proof as an enumerable sequence of *A***PredCal**-formulas each of which is either an *A***PredCal**-axiom or arises therefrom by finitely many applications of rules $T1$ and $T2$; an *A***PredCal**-theorem as the terminal member of any A**PredCal**-proof; and an *A***PredCal***-formula's* Φ *being deducible from* a set Φ of *A***PredCal**-formulas as Φ's being the terminal member of an *A***PredCal**-sequence of formulas each member of which is either an *A***PredCal**-axiom or a member of Γ or arises from preceding formulas of the sequence by finitely many applications of $T1$ and $T2$.

[13]If we are only concerned with *proofs* (premiss-free deductions, deductions from axioms) there are no provisos. If we want to admit deductions from ad hoc premiss, a sufficient proviso would be that *if* α does not occur free in Φ *then* α does not occur free in any of the (ad hoc) premises of the deduction as a whole. But this proviso can be liberalized to refer only to those premisses on which the occurrence of Φ used in the deduction step actually depends. (It's a bit of work to define this exactly.)

A**PredCal** exhibits some peculiarities about which a word of explanation is needed. For one thing, none of $A1$ to $A5$ is itself a **PredCal**-formula; in fact they are not even written in the language of **PredCal**. We here chose to use our metalanguage to specify (though not actually record) an infinite number of axioms for **PredCal**; thus each of the As is an axiom *schema*, written in the metalanguage, that specifies as axioms an infinite assortment of **PredCal**-formulas having the syntactic form of the schema.

Secondly, it may not be clear why we trouble to impose restrictions on axiom schemas $A4$, $A5$ and rule $T2$. From a purely formal point of view, there is no particular reason for it, but none against it either. However, a major part of the *motivation* of an axiomatization of any system is the desire to generate as uninterpreted theorems exactly those formulas of the system that are there valid, i.e. true for every model/structure. In the present case, the restrictions are imposed to avoid the necessity of countenancing as axioms of A**PredCal** (and hence as theorems of A**PredCal**) formulas that are not **PredCal** valid. Thus, in $A4$, unless we had the restriction, we could have the following situation. Suppose that there exists an infinite domain, D, of towns no one of which has the same official population as any other. Let D be the domain of an interpretation of A**PredCal**; let "F_1^2" represent the predicate, "has the same official population as". Then, under this construal, we would have the sentence

(1) $(\forall x_1 \neg \forall x_2 F_1^2(x_1, x_2)) \supset \neg(\forall x_2(F_1^2(x_2, x_2))),$

represent the (clearly false) English sentence

($1E$) If no town has the same official population as every town, then not every town has the same population as itself.

Now, by $A4$ *minus* the restriction, (1) is an axiom. Yet (1) is plainly invalid for our structure, since the antecedent holds for that structure and its consequent fails. The restriction performs the needed job of forbidding the substitution on (1) of an occurrence of "x_2" for an occurrence of "x_1", since "x_2" is not free for "x_1" in (1). (As without restriction would make an axiom of the invalid formula $\forall x(Fx \forall Fx) \rightarrow (Fx \rightarrow \forall x Fx)$.)

A similar story can be told of the restriction in T2. For, again using the above domain of interpretation, let "F_1^2" represent the English predicate "is home of the Rimrock Hotel". Then the sequence

(2) $\langle (F_1^2(x_1))(\forall x_1(F_1^2))\rangle$

would exemplify a good **PredCal**-deduction of $\forall x_1 F_1^2 x$ from $F_1^2 x$, even though the sequence,

$\langle$Banff, Banff, …, Banff, …$\rangle$

doesn't satisfy $\forall x F_1^2 x_1$. For, understanding Banff to be the first element of this sequence, "$F_1^2(x_1)$" is satisfied by that sequence, Banff being in fact the home of the Rimrock Hotel; but "$\forall x_1(F_1^2)$" is not satisfied by it since it is not true that the sequence,

$$\langle \text{Standoff, Banff, Banff, } \ldots, \text{ Banff, } \ldots \rangle$$

which differs form the original sequence only at the kth (here, the first) element, satisfies the formula. Thus, for that structure, $\forall x_1 F_1^2$ is not satisfied, and therefore the deduction cannot be admitted from the point of view of structures.

We conclude the present section with a selection of theorems of *A***PredCal**. They are specified by means of the following theorem schemata, which stand to axiom-schemata as theorems stand to axioms.

*Some Theorem Schemata of A***PredCal**

Th(1): $\ulcorner\forall\alpha(\Phi) \supset \forall\beta(S^\alpha_\beta\Phi)\urcorner$, provided β does not occur free in Φ and β is free for α in Φ.

Th(2): $\ulcorner(\forall\alpha(\Phi)) \equiv \forall\beta(S^\alpha_\beta(\Phi))\urcorner$, provided β does not occur in Φ and β free for d in Φ.

Th(3): $\ulcorner(\forall\alpha(\forall\beta(\Phi))) \equiv (\forall\beta(\forall\alpha(\Phi)))\urcorner$.

Th(4): $\ulcorner(\Phi \supset \forall\alpha\Psi) \supset \forall\alpha(\Phi) \supset \Psi)\urcorner$, provided α is not free in Φ.

Metatheory of *A*PredCal. All this, of course, prompts one to ask, in particular, whether *A***PredCal** manages to produce as theorems just all the valid *A***PredCal** formulas; and, more generally, whether *A***PredCal** has other interesting properties, and if so what are they?

To the first question, the answer is yes. *A***PredCal** is *correct* and in the sense that exactly its theorems are exactly its valid formulas.[14] (Proved by Gödel [1930] and later by Henkin [1949]). The metaproof is here omitted.

To the second question, the answer is also yes. *A***PredCal** is *consistent* in the sense that for no formula Φ is it the case both that Φ is a theorem and also that $\ulcorner\neg\Phi\urcorner$ is a theorem.

That *A***PredCal** is consistent is important, since every consistent such theory has a model. (Proved by Gödel [1930]). Thus *A***PredCal** has a model in fact every model is a model of the set of formulas of *A***PredCal**. *A***PredCal** Detachment and Generalization are "good" Transformation Rules, since under them *sequence-satisfaction*, *truth-in-a-structure*, satisfaction by all sequences that satisfy the premises of the deduction, and *validity* are hereditary.

[14]Note that the soundness metatheorem for **PC** gives one-half of the completeness property in the the present meaning of "completeness".

A**PredCal** admits of a *Compactness Metatheorem* (proved by Church [1936]. Also provable is a *Deduction Metatheorem.* As we saw in our discussion of **PC**, the Deduction Metatheorem is a useful device for discovering theorems, especially if theoremhood is not a decidable property. In **PC** theoremhood is decidable, since validity is, and **PC** is correct and complete. But in A**PredCal** neither validity nor theoremhood is decidable. (Proved by Gödel [1930]). However a fragment of A**PredCal** is decidable. Let (A^1**PredCal**) be exactly like A**PredCal** save that all its predicates are one-place (they are the so-called *monadic* predicates). A^1**PredCal** is "the monadic predicate calculus". Validity is decidable in A^1**PredCal**. This last is important if only in that it suggests that the richer the syntax of a formal system the less likely its central notions are to behave algorithmically. And this is so.

9 ENGLISH AND **PREDCAL**

When we move from **PC** to **PredCal** this happy convergence *seems* to fall apart. Quantifiers are the problem. The problem they give for convergence is that they are not (much of) a problem for proof theorists but are for theorists of the valid-argument persuasion. And for the model theorist they are a very hard problem indeed. The argument theorist is faced with the difficulty that the quantifiers of **PredCal** seem not to have the same grammar as those of English. In other words, the **PredCal** quantifiers are counterintuitive. It is worse for the model theorist. Let

$$\Phi_1, \ldots, \Phi_n, \Psi$$

now be quantified sentences of **PredCal**. If the Φ_i together logically imply Ψ then there is no structure in which the Φ_i are true and Ψ false. But what is a structure? It is the natural child of algebraic parents, an abstract set-theoretic object. Beyond that there are a great many of them, certainly more than denumerably many. Abstract and ubiquitous, they are everywhere around but awfully hard to recognize. So how can it be determined whether Ψ is true in *every* model in which $\Phi_1, \ldots, \Phi_n$ are also true? The model theorist must bring models to heel. It is customary to suppose that he does this by invoking the constraints of the cumulative hierarchy of *sets*, as provided for in (say) the Zermelo–Fraenkel (ZF) axioms. By these lights V is the universe of sets and is itself a structure. In a crucial respect V is thought to be on a par with lesser sets: not only are the ZF axioms true in them, so too the sentences of **PredCal** are true or false in them. In fact, however, there is no notion of truth that is both applicable to V and formally articulable in ZF. Strictly speaking, then, the model theorist can't have 'true' in the way that he aspires to have it as a logician i.e., as the output of his own theory; (so he can't have 'all' and 'some' in that way

either). In this respect, the model theorist is in the same boat as the **PC**-representer of English on the hoof. He wanted his judgements of validity and consistency and invalidity to be secured by logical forms. For the last two, there is no formal way to satisfy **Logical Inertia**, and even if it is satisfied *somehow*, the **PC**-logician has lost consistency and invalidity in the way that he contrived to have them in the first place. If the appearances of late count for anything, the argumentation theorist comes second in the **PredCal** stakes. For fanatical traffickers after a theory of natural language argument this is occasion to abjure **PredCal**. For the less excitable, there is reason to consider a marriage of convenience with proof theory. It is of no use. In **PredCal** all three approaches come to the same thing. Convergence on how to handle formulas of the form $\ulcorner\Phi_1, \ldots, \Phi_n\urcorner$ logically implies Ψ is restorable after all (Hodges [1983], section 17).

Still, quantifiers in English are hard to make out. It is galling that two such little words, 'every' and 'some' should act up so. But they do. What is the problem? Let us attend to Michael Dummett [1973, 517]:

> ... given that we understand a sentence from which a predicate has been formed by omission of certain occurrences of a name, we are capable of recognizing what concept that predicate stand for in the sense of knowing what it is for it to be true of or false of an arbitrary object, whether or not the language contains a name for that object.

Dummett catches an essential technical point for the development of **PredCal**. It is this: that if "Mary is angry" is a sentence and, so, eligible for truth functional combination, then "x is angry" is a sentence-*like* construct and hence also eligible for truth-functional combination. By extension, "Every woman is angry" involves a quantifier expression and two predicates, or as we may now say, *open-sentences*: 'x is a woman' and 'x is angry', where 'x is a woman' is a predicate. Predicates are open sentences; they can be truth functionally combined, as witness

x is a woman $\supset$ x is angry.

'Every' now is 'every x' and 'every x' is qualified by the relative predicate 'x is a woman'. The second predicate 'x is angry' qualifies what went before. Thus, as reconstructed in **PredCal**, "Every woman is angry" is

For every x, x is a women $\supset$ x is angry

or, in a stylistic variation,

$\forall x$ (x is a woman $\supset$ x is angry).

We said that quantifiers were difficult. Part of their difficulty is now overcome. We overcame it by seeing that predicates accessible to quantification

are *sentences*. It is something that Peirce [1885] was exactly right to emphasize.

The quantificational expression 'some' is given like treatment. "Some women are angry" is construed as

For some x, x is a woman $\wedge$ x is angry

or, in a notational variant,

$\exists x$(x is a women $\wedge$ x is angry).

Here too the sentence is parsed into two expression types: a quantificational expression, '$\exists x$', and the open-sentences, 'x is a woman', and 'x is angry'.

English abounds in quantifier-expressions — 'most', 'many', 'few', 'a few', 'several', etc.— that cannot be reconciled to the grammar of 'all' and 'some', as has been demonstrated repeatedly. (Barwise and Cooper [1981]). This answers a question that bothers some critics of logic as a theory of natural language: Why does **PredCal** restrict itself to just two quantifiers? The short answer is that they are the only two that fit the grammatical paradigm at hand.

Even so, it is clear that **PredCal** affords us the opportunity of representing very many more kinds of English arguments than did **PC**, and many more kinds of English sentence than it. But it would still be a mistake to think that the language of **PredCal** is capable of reflecting all the logically relevant syntactic structure of English. For example, **PredCal** has no *identity predicate*, and so cannot (tightly) represent such inferences as:

1. Cataline opposed Cicero
2. Cicero = Tully.
3. So, Cataline opposed Tully.[15]

Neither does **PredCal** contain any *function symbols*, nor, other than the connectives, operator symbols, and so cannot (tightly) represent such truths as:

4. *The sum of 2 and 4* is 6.
5. *4 x 6* = 24
6. If *necessarily* man is mortal then man is mortal.

PredCal also does not acknowledge tenses and adverbs of tense, so cannot (tightly) represent:

7. If Sam *was* here, then possibly he is not here *now*.

How far should we be prepared to go in our efforts to reflect the logical structure of English?

[15]This is disputed by Quine, where Quine [1970, chapter 5] suggests the contextual eliminability of "=" from otherwise standard **PredCal** languages.

10 FIRST ORDER LOGIC

The language of **PredCal** can be extended to include the identity predicate, and some function symbols. And after a few new axioms are added, the result is a system that can handle examples of the sort (1) to (5). Roughly, such a system is said to be an *elementary first order quantification theory*, *F***PredCal**, adequate for the expression of ordinary arithmetic. Whether to go further than *F***PredCal**, adding the vocabulary and the axioms needed to represent cases (6) and (7), is currently a matter of lively controversy. The conservatives have a point when they say that such non-standard systems are (comparatively) not very well known, are not yet in many cases very deeply studied, so why buy a pig in a poke? The non-standardists have a point when they remind us that *F***PredCal** is not complete, (proved by Gödel [1931]). Note that *F***PredCal** (the logic of first order with identity and function symbols) itself *is* complete given the standard axiomatization (just as *A***PredCal**). Gödel established that any consistent first order theory *T*, if it is equipped with an effective procedure for determining whether an object is one of its own proofs, is incomplete. That is, if it has the predicate 'is a proof of *T*' the class of objects of which that predicate is true is a decidable class, all recursive functions are representable in *T* (so that the fixed point theorem maybe proved), *F***PredCal** is incomplete. So it is not much of an argument on *F***PredCal**'s behalf that many of these more elaborate non-first order systems are not, or are not known to be, complete. How *much* of a point either has plainly depends on what one supposes an adequate logic to be. For excellent discussions of first order theories see Goldfarb [1979], Moore [1980], Hodges [1983, sections 13–15, 26], and Barwise [1977].

11 VARIATIONS ON STANDARD SYSTEMS

Arguers have a standing interest in two questions. One is whether an argument that someone has already *made* is a good argument. The other is how to *go about* making a good argument. In a streamlined version, the second is a question about the conditions under which a search for a good argument is made *efficiently*. Belief-revision is likewise something to which both these questions naturally apply. As we saw, one attraction of syllogisms is their inchoate recognition of the importance of the second question. If, in a search for a syllogism for a conclusion *C* the enquirer were free to deploy any premiss that left intact the relation of entailment between all the premisses he has so far considered and *C*, then any enquirer exercising this freedom unconstrained would be a radically inefficient syllogizer. Better if he had some way of discounting large classes of useless premisses. But Aristotle's non-superfluousness condition on premisses is not in fact a satisfactory condition on premiss searches. A premiss cannot be recognized as superfluous in

Aristotle's sense unless it occurs in a valid argument. But where the search is for valid arguments, Aristotle's condition won't help. It merely tells you what to do with useless premisses after you have succeededin making a valid argument out of them.

Certain modern systems of logic try to go Aristotle one better in the business of premiss searches. They are not standard systems, but rather are restrictions of **PredCal**. Perhaps the best-known nonstandard system having efficiency of premiss search as a basic motivation is the system R of *relevant propositional logic*, developed by Alan Ross Anderson and Nuel D. Belnap, Jr. [1975] and Anderson *et al.* [1992], and the veritable galaxy of extensions, enrichments and variations that orbit R. (Routley *et al.* [1982]).

Efficiency of premiss-search is a matter of subduing complexity. Complexity is a *computational* property. Thus a search for efficient premisses is a quest for a computational simplicity. At the propositional level, **PC** properly contains *intuitionistic* (propositional) *logic*, which in turn properly contains *minimal* (propositional) *logic*. (Prawitz [1975]). They can be distinguished by reference to their respective policies for handling hypotheses in the form $\{\Phi, \neg\Phi\}$. In **PC** all rules mentioned are valid and so is *classical* reductio, and any formula is derivable from $\{\Phi, \neg\Phi\}$ (*ex falso sequitur quodlibet*) Int.$\rightarrow$IPC lacks: double negation, excluded middle and *classical* reductio (with regular reductio retained), any formula still derivable from $\{\Phi, \neg\Phi\}$ constructive dilemma is valid in all three systems; (disjunctive syllogisms valid in **PC** and intuitionistic logic but not in minimal logic) minimal$\rightarrow$MPC lacks the property that *any* formula follows from $\{\Phi, \neg\Phi\}$, but all negations still follow.

Of the three, **PC** is a comparative computational delight. **PC** is *NP-complete*. (Cook [1971]). The problem of whether a sentence Ψ has a proof in **PC** belongs to a class K of computational problems. K is NP-complete just in case two conditions are met. First, K-problems can be solved in nondeterministic polynomial time (in which case K is *NP-easy*). Second, any other solvable problem class K^* can be transformed into K in deterministic polynomial time (hence, is *NP*-hard).

A K is solvable in deterministic polynomial time if there is a polynomial function Φ such that any problem in K of length n is solvable in $\Phi(n)$ units of time.

A K is solvable in nondeterministic polynomial time if and only if there is an effective branching procedure β for solving it and a polynomial function Φ such that for any argument A to Φ of length n, then the tree generated by β contains a solution of A at the termination of *some* branch of length less than $\Phi(n)$. Thus if the user of β were well-favoured enough to "hit upon" the "right" branch, he would find the quickest solution, a solution guaranteed to occupy him for no longer than the $\Phi(n)$ units of time. *If*, indeed!

> *NP* is the class of all K computable in nondeterministic time. P is the class of all classes K computable in (deterministic) polynomial time.

No one has produced for the thousands of Ks known to be *NP*-complete an effective procedure that runs in polynomial time. Thus the received wisdom is the $P \neq NP$, even though the question is strictly open.

PC has a whole bag of proof strategies. So it is not surprising that, with all those resources, it is *N*P-complete. Intuitionistic logic and minimal logic don't do nearly as well. They are both very hard, i.e., PSPACE-complete.[16]

That **PC** is *NP*-complete and the others not should give pause to any theorist of argument or of inference looking for a tractable model of how things are on the hoof. In particular, anyone drawn to relevant systems such as R out of an interest in computational efficiency will need to deal with the possibility that **PC** sets a benchmark for tractability. Let us see.

PC is decidable. R is not (Urquhart [1984]). *LR*, a fragment of R got by removing the distributivity axiom, is decidable (Thistlewaite, McRobbie, and Meyer [1988]), but its decision problem is dauntingly inefficient. *LR* is no better than ESPACE hard, and may be exponential space-hard[17] on the generalized Ackermann's function. (Ackermann's function dominates any primitive recursive function. See Hilbert and Ackermann [1928], and Urquhart [1990] and [1999].)

One of the reasons that Anderson and Belnap had for trying to bring relevance within the orbit of an otherwise classical logic was to make proof searches more efficient. For this to happen, R's set of search problems would need to be *P*-complete (on the assumption that $P \neq NP$), since **PC**'s search problems are *NP*-complete. The analytic apparatus with which Anderson and Belnap attempted to corral the relevance relation has three main components. One is that the metatheoretic predicate "is a proof of ..." is a transitive relation. Another is that in R and its satellite systems the classical rule of disjunctive syllogisms fails. The third innovation is that the connectives in R and its satellites are at least partially intensional (or, *non*-truth-functional). In fact, although not intended by its originators, it is provable that in the negation-implication fragment of R, *all* connectives must be intensional if any is (Avron [1992]).

[16]PSPACE is the set of decision problems determinable by polynomial space-bounded deterministic Turing programs that halt for any input. A decision class K is PSPACE-complete if ϵ PSPACE, K^* polynomially implies K. So even if it were the case that $P = NP$, it is possible that $P \neq$ PSPACE. (Garey and Johnson [1979, 171]). A problem is PSPACE-hard if some recognized PSPACE-complete problem π is Turing reducible to it without the need to state that π itself is in PSPACE.

[17]A decision problem is (exponential) space-hard if and only if a Turing machine on which the problem is solvable employs exponential quantities of (decision) space on infinitely many inputs. The implication fragment, $R \rightarrow$, of R is space-hard.

It is striking and ironic that in systems such as R the analysis of relevance itself should be the very thing that makes them irredeemably hopeless from the point of view of computational efficiency. It is a salutary lesson: *Any* restriction worthy of the name in **PC** will make for greater not less intractability.

Of course, computational efficiency is not an issue when a system of logic does not purport to represent or model how human reasoners actually produce truth-preserving belief-revisions. This leaves it open that comparatively intractable systems might appeal for other reasons. One is that there really *is* a relevance constraint on the entailment relation, and that it is a bad thing, therefore, that systems such as **PC** and **PredCal** leave it unapplied (and unrecognized). This is not the place to decide such issues. But we mention in passing a system of relevant propositional logic (which has a natural quantificational extension) in which consideration of the "true nature" of entailment and of the complexity of searches for its instances come into happier alignment than Anderson–Belnap systems could ever allow. The system is Neil Tennant's *IR* of relevant intuitionistic logic. (Tennant [1987, part II], [1992].) *IR* is a minimal logic augmented by disjunctive syllogism and the rule

$$\Phi \vee (\Psi \wedge \neg\Psi) \vdash \Phi.$$

In *IR* the transitivity of proof is not unrestrictedly valid. Tennant likes *IR* because it "has strong meaning-theoretic, epistemological and methodological claims to adequacy and correctness" (Tennant [1992, 6]). For example, he says that it is "adequate for Popperian science" (Tennant [1987, 201 ff]). How does *IR* fare on the question of complexity? It does better than *R*, since it is no less tractable than intuitionistic logic, that is, it no more decisionally complex than PSPACE-complete. On the other hand, Tennant is also able to show that a *non*-Anderson-Belnap system *CR* of relevant propositional logic is constructible from PC and that its decisions problems are no more complex than those of **PC**, which is *NP*-complete.

12 ARGUMENT AND INFERENCE

We have already seen that no standard system of logic stand as a satisfactory theory of deductively appraisable arguments on the hoof. Their utility in this regard depends essentially on a representation policy *R* which takes stretches of natural language discourse into grammatical entities constructed out of the language of some or other deductive system. Among argument theorists there are critics without number who object that any standard *R* strips its natural language inputs of a hefty abundance of its actual properties, whether contextual, syntactic or semantic. This is right but is nothing to regret as such. What matters are the argumentation theorist's target properties and whether the logician's representation policy has

the backwards reflection property with respect to those targets. If the argumentation theorist has no interest in whether an argument on the hoof possesses context-insensitive properties or properties that turn on syntactic and semantic features which the representation procedures can recognize, then logic has nothing to say about what interests him about arguments on the hoof. But what an argumentation theorist may be interested in does not fix the properties that arguments on the hoof actually have. Of course, as we have said, *tight* representation policies require their inputs to satisfy the Logical Inertia Principle, and whether they do or not, no standard system of formal logic is able to determine. But that this is so leaves the logician and the argumentation theorist at parity. Neither has a good theory of logical inertia for natural languages, and both would be well-served by having one.

Then, too, branching quantification is a problem. For

(1) Some melody of every big band from the forties is copied in some hit song of every rock 'n roll star (Hacking [1979, 308]).[18]

More generally, if we try to say that for each x there is a y and for each z there is a w such that Fx, y, z, w, how would we represent this in **PredCal**? We could try

2. $\forall x \exists y \forall z \exists w (Fxyzw)$

or

3. $\forall z \exists w \forall x \exists y (Fxyzw)$

It won't work. Whereas we want the choice of y to depend only on x and the choice of w to depend only on z, in (2) "we represent the choice of w as depending on x too ...", and in (3) "we represent the choice of y as depending on z, too" (Quine [1970, 89–90]).

"As a way of avoiding these unwanted dependencies" (Quine [1970, 90]), we might try

4. $\forall x \exists y$
   $Fxyzw$
   $\forall z \exists w$

which is a *branching* structure equivalent to neither (2) nor (3). But (4) is not a formula of **PredCal**, and a good thing too some will say. For "as soon as you branch out in the manner of (4) you get into a terrain that does not admit simultaneously of complete proof procedures for validity and consistency" (Quine [1970, 91]; proved by Henkin [1961, 181] and Craig [1957, 281]). Quine points out that our branching quantificational utterance

[18]Hacking attributes the example to Janet Behner and cites Carlsen and ter Meulen [1979].

has a neat representation in a second order quantification theory, that is, a many-sorted system in which there is quantification over a distinguished set of non-individuals or, as in this case, functions. Thus

1. $\exists f \exists g \forall x \forall z (F(x, fx, z, gz))$.

Concerning (1), Hacking is of the view that there is nothing to be found that couldn't be captured in the manner of (5) and that its functional expressions 'f' and 'g' can be restricted "to very low-grade ramified functions." [Hacking, 1979, 308]. Quine and Hacking differ as to whether (5) is a sentence of a *logic*, with Hacking giving it the nod and Quine demurring. There is no doubt, however, that (5) itself is not constructable in **PredCal** or in any standard first order theory, and it is this that the argumentation theorist may want to make something of. Indeed he has a new point to make: There are quantified *syntactic* structures in English that won't go over into the language of any standard predicate calculus, and that this is so doesn't turn on the failure of any semantic principle such as Logical Inertia.

The heart and soul of formal deductive logic is the citation of target properties and the specification of logical forms such that those properties are intrinsic to those forms. Of central importance is the operation of substitution, which is a process that also preserves logical form under syntactic variation, and also preserves the formal target properties under such variation. Languages for which substitution is universally target-property preserving are called "extensional". In the case of the object languages of systems such as **PC** and **PredCal**, target properties are closed under substitution, and so such languages are "extensional". In the case of natural languages it is clear that they are not extensional just as they come. There are natural language structures galore whose target properties are not possessed by their representations in any extensional object language. This calls for hefty constraints on the logician's formalization policies. He must try to contrive his representational strategies in ways that honour the backwards reflection principle with respect to the theory's target properties. For this to happen, he must restrict the inputs to the formalization function in ways that enable his theoretical objectives to be met. This amounts to a restriction on inputs which limits them to those natural language structures with respect to which the theorist's target properties hold of them in virtue of their logical forms. These are weighty restrictions. They deny to the input-position natural language entities which fail Logical Inertia, which involve branching quantifiers, which admit unrestricted adjectival and adverbial modification, which display functional expressions, or certain various predicate constants, such as existence. This leaves extensional logic with at most a structurally impoverished proper subset of any natural language on which it has any claim to pronounce; and it leaves the critical argumentation theorist with three complaints to press.

First complaint: Since most good real-life arguments are judged bad (i.e., invalid) by any extensional logic purporting to represent them, extensional logic has little to offer a theory of arguments on the hoof.

Second complaint: Owing to the harsh requirements of extensionality, large classes of arguments on the hoof that actually possess target properties such as validity, are arguments, having no representations in standard logical systems.

Third complaint: The logician has no wholly satisfactory principled basis on which to distinguish between those natural language structures on which his logic is fitted to pronounce and those on which it is not fitted to pronounce.

There are contemporary logicians who are more than prepared to throw in the towel on the question of logic's fruitfulness as a theory of arguments on the hoof. Quine [1970] is perhaps the most emphatic of these. Extensional or classical logic is not a theory of arguments on the hoof and not a theory, either, of ordinary inference. Better, says Quine, to regard first order logic as a filter on natural language whose output suffices, when supplemented by the appropriate empirical predicates, for the exact formulation of scientific theories. Central to such a suggestion is the claim that extensionalized English suffices for the expression, in principle, of any statement of a scientific theory. Such a claim, whatever its initial appeal, is done considerable damage, however, by the broad recognition that quantum physics is almost certainly not satisfactorily formulable in a purely extensional language.

The dominance of classical logic granted, some of the most impressive developments in the present century have occurred in the area of *intensional* logic. Here we meet with calculi for alethic modalities ("necessarily" and "possibly"), causal modalities ("causally necessary", "causally possible"), temporal quantifiers, logic of tense, logic of events, and relevant logics. Also of note are systems of dynamic logic and situational logic in which the meaning of logical particles is not fixed by context-free truth conditions on sentences but in which reasoning is represented as information processing under contextually sensitive constraints. It would be too much to say that all such systems are motivated by the desire to contrive more accessible and realistic theories of arguments on the hoof and of ordinary inference and practical reasoning. Even so, it is possible virtually without exception to see in these theories *something* of such a motivation. Their success or failure in such regard is a matter on which the jury is still out and, in any event, is a matter that exceeds the reach of the present chapter. Still, these are developments for the argumentation theorist and the theorist of practical reasoning to keep an eye on. For the central fact is that natural languages are intensional. How, then, can it be realistic to look for its logic in purely extensional ways?

This is a fact of some importance. So we shall close the present chapter with a brief sketch of intensional logic, in the manner of Gamut [1991].

13 INTENSIONAL LOGIC

The intensionality of an intensional logic is a matter of degree. It is a matter of its degree of departure from a *purely extensional logic.* A logic is purely extensional if it satisfies the following conditions.

1. Its connectives are truth-functional.
2. Substitution of co-denoting terms preserves singular reference in all referential contexts.
3. Substitution of co-designating predicates preserves predicate-designation in all predicative contexts.
4. Substitution of co-valued sentences preserves truth in all sentential contexts.

To save space, we shall confine our attention to propositional logic. The most developed intensional approaches to propositional (and predicate) logic involve the addition of sentence operators (one or more), which intuitively convey the idea of

necessity and possibility (modal logic)
tense and/or time (tense/and or temporal logic)
knowledge and belief (epistemic and doxastic logic)
permission and obligation (deontic logic).

We shall here examine both the modal and tensed/temporal approaches to propositional logic. Common to all such logics is the idea that sentences are assigned truth-values relative to some specified context c. Let C be the set of such contexts. Then the truth of a sentence in context c will depend on its truth in c but *also* in other contexts $c*$ in C. Let c and $c*$ be points of time. Then the sentence ⌜It was once the case that Φ⌝ is true at c, only if there was an earlier time $c*$ at which Φ itself is true. Similarly, "It is possibly the case that Φ" is true in the actual world c only if there is some suitably alternative world $c*$ in which Φ itself is true.

Modal propositional logic **MPC**

To the language of **PC** we add the *modal operators*indexmodal operators $\Box$ and $\Diamond$, and the new Formation Rule:

i) If Φ is a wff of **MPC** then $\ulcorner\Box\Phi\urcorner$ and $\urcorner\Diamond\Phi\urcorner$ are so as well.

By the provisions of i), the following are examples of **MPC**-formulas.

$\Box p$
$\Diamond p \vee \Box q$
$\neg\Box(p \wedge q)$
$p \supset \Box\Diamond p$
$\Box \supset \Diamond\Box p$

In the semantics of **MPC**, there are two fundamental intuitions. One is that the truth of $\ulcorner\Box\Phi\urcorner$ and $\ulcorner\Diamond\Phi\urcorner$ in any given world will be a matter of Φ's truth in other possible worlds.

We now introduce the idea of a model for **MPC**.

ii) A *model* M for **MPC** is a non-empty set W of possible worlds, a binary accessibility relation R on W, and a valuation V which assigns a truth value $V_w\Phi$ to every atomic sentence in each world $w \in W$.

The pair $\{W, R\}$ is also called a *structure*. Hence M is a structure together with a valuation.

The truth value of a sentence Φ in a world w in model M ($V_{M,W}\Phi$) is defined as follows

iii) $V_{M,W}\Phi = V_W\Phi$, for all atomic wffs

iv) $V_{M,W}\neg\Phi = T$ iff $V_{M,W}\Phi = F$

v) $V_{M,W}(\Phi \supset \Phi) = T$ iff $V_{M,W}\Phi = F$ or $V_{M,W}\Psi = T$

vi) $V_{M,W}\Box\Phi = T$ iff for all $w* \in W$ such that $wRw* : V_{M,W}\Phi = T$

Thus "necessarily true" means "true in all accessible worlds", and "possibly true" means "true in at least one accessible world".

For any model M it is always some wffs that are true in each of M's possible worlds. Two examples are

$\Diamond p \wedge \Diamond\neg p$
$\Box p \supset p$

Formulas true in all worlds of a model are valid in that model. If Φ is such a sentence, its validity is expressed as

$V_M\Phi = T$

The class of sentences valid in a model M subdivide into those whose validity is dependent on the particular valuation that M has, and those whose vlaidity is independent of such considerations. These latter sentences owe their validity soley to M's structure, i.e., to its set of worlds W and its accessibiltiy relation R. "$\Box p \supset p$" is an example of such a formula. We can now say that a wff Φ which is valid in every model just on this basis of its structure S, then Φ is valid on S. It can be shown that "$\Box p \supset p$" is valid on any structure whose accessibility relation is *reflexive*.

Modal logicians are especially interested in exposing in detail relations between the validity of wffs and various properties of structures. Consider for example the schema

$\Box(\Phi \supset \Psi) \supset (\Box\Phi \supset \Box\Psi)$.

As it happens it is valid in every structure *independently* of the properties of the accessibility relation R. On the other hand, the validity of the schema

$$\Box\Phi \supset \Box\Box\Phi$$

depends on the *transitivity* of R. And the schema

$$\Diamond\Box p \supset \Phi$$

owes its validity to R's *symmetry*.

We now turn to a propositional tense logic (**TPC**). Tense logic is notationally similar to modal logic. In tense logic, contexts are not worlds but rather moments of time; (or, in some approaches intervals of time), and the accessibility relation is that of earlier than.

TPC results from **PC** by addition of the operators G and H, each similar to $\Box$. G is read as "it is always going to be the case that" and H as "it always has been the case that". Corresponding to $\Diamond$ are the further two operators (definable from the others) F and P, which read respectively as, "will in the future be the case that" and "was in the past the case that".

Let us consider Φ, the atomic sentnece "Harry is rich". Then **TPC** has the means of recognizing the following tenses

Φ	Harry is rich
$F\Phi$	Harry will be rich
$P\Phi$	Harry was rich
$PP\Phi$	Harry had been rich
$FP\Phi$	Harry will have been rich
$PF\Phi$	Harry would be rich

(Note, however, that the distinction between the simple past and present perfect can't be recognized in **TPC**.)

A model M for **TPC** is a nonempty set T of moments in time, an "earlier than" relation R, and a valuation V, which for each atomic wff Φ and each moment $t \in T$ assigns a truth value $V_t\Phi$ to Φ at t. Here, too, T and R define a structure, which is also sometimes called a *time axis*.

Now comes a truth definition for the various tenses.

i) $V_{M,t}G\Phi = T$ iff for all $t* \in T$ such that $tRt* : V_{M,t*}\Phi = T$

ii) $V_{M,t}F\Phi = T$ iff for at least one $t* \in T$ such that $tRt* : V_{M,t*}\Phi = T$

iii) $V_{M,t}H\Phi = T$ iff for all $t* \in T$ such that $t*Rt : V_{M,t*}\Phi := T$

iv) $V_{M,t}P\Phi = T$ iff for at least one $t* \in T$ such that $t*Rt : V_{M,t*}\Phi = T$.

As with modal logic, there are principles of tense-logic which express (depend on) certain properties of structures (or time axis). For example,

$$G(\Phi \supset \Psi) \supset (G\Phi \supset G\Psi)$$

and

$$H(\Phi \supset \Psi) \supset (H\Phi \supset H\Psi)$$

are valid independently of the accessibility relation, hence valid on any time-axis. (Recall, that these are analogues of $\ulcorner\Box(\Phi \supset \Psi) \supset (\Box\Phi \supset \Box\Psi)\urcorner$, which is likewise valid in every structure.) Unlike $\ulcorner\Box\Phi \supset \Phi\urcorner$ the next two analogues are not wholly intuitive (but see below).

$$G\Phi \supset \Phi$$

and

$$H\Phi \supset \Phi$$

would be invalid. Lacking a formula to express this irreflexivity, the two formulas at hand survive as principles of **TPC**, as do:

v) $\Phi \supset HF\Phi$

vi) $\Phi \supset GP\Phi$

vii) $P\Phi \supset H(F\Phi \vee \Phi \vee P\Phi)$

viii) $F\Phi \supset G(P\Phi \vee \Phi \vee F\Phi)$

ix) $P\Phi \supset GP\Phi$

x) $F\Phi \supset HF\Phi$

(v) says that what is now so has in the past always been something that would occur. (vi) provides that what is now so will always be something that has occurred. (vii) syas that if Φ was once so, then it has always been that either Φ was yet to occur or that Φ was occurring or that Φ had already happened.

Principles (v) and (vi) are true on every time-axis. Principles (vii) and (viii) are valid on all time-axes where R is *connected* (i.e., such that for any two times, one is earlier than the other).

Schemata (ix) and (x) are valid iff R is *transitive*.

Provided that time is a *linear order* (i.e., transitive, asymmetric, and connected), then all the above schemata of **TPL** are valid.

We will say that R is *dense* if it satisfies the condition:

> If for all distinct x and y x bears R to y, then there is a distinct z such that x bears R to z and z bears R to y.

Given that R is dense, then there are three more valid schemata.

xi) $F\Phi \supset FF\Phi$

xii) $\neg G(\Phi \wedge \neg\Phi)$

xiii) $\neg H(\Phi \wedge \neg\Phi)$.

From the elementary discussion of the present section, it is easy to see that intensional logics, even at the propositional level, are interpretable in ways that make their languages far better simulacra of actual languages than anything to be found in a strictly extensional language. That is clearly a gain for anyone interested in a realistic logic of natural language argument and inference. But, as we saw with standard systems of relevant logic, intensional logics are computationally daunting. Complexity is a serious problem for extensional logic. It is a much more serious problem for intensional logics. It is necessary to ask whether computational costs are a reasonable price to pay for the benefits that accrue to linguistic richness. As we shall see, this is also a problem for inductive.

ACKNOWLEDGEMENT

I am greatly indebted to Professor Erik C. W. Krabbe for detailed and insightful suggestions, on a scale that surely must set new standards for collegial generosity.

BIBLIOGRAPHY

[Anderson and Belnap, 1975] Alan Ross Anderson and Nuel D. Belnap. *Entailment: The Logic of Relevance and Necessity*, volume 1. Princeton, NJ: Princeton University Press, 1975.

[Anderson *et al.*, 1992] A. R. Anderson, N. D. Belnap Jr., and J. M. Dunn. *Entailment*, volume 2. Princeton, NJ: Princeton University Press, 1992.

[Avron, 1992] Arnon Avron. Whither relevance logic? *Journal of Philosophical Logic*, 21:243–281, 1992.

[Barnes, 1994] Jonathan Barnes. *Aristotle's Posterior Analytics*. Oxford: Clarendon Press, 1994. 2nd edition.

[Barwise and Cooper, 1981] Jon Barwise and R. Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4, 1981.

[Barwise, 1977] Jon Barwise. On Moschovakis closure ordinals. *Journal of Symbolic Logic*, 42:292–296, 1977.

[Bell and Machover, 1977] J.L. Bell and M. Machover. *A Course in Mathematical Logic*. Amsterdam: North-Holland, 1977.

[Beth, 1955] Evert W. Beth. Semantic entailment and formal derivability. *Mededelingen der Koninklikjke Nederlandse Akademie van Wetenschappen (Amsterdam)*, 18:309–342, 1955. Afd. letterkunde, n.s.

[Carlsen and ter Meulen, 1979] Laurie Carlsen and Alice ter Meulen. Informational independence in intensional contexts. In E. Saarinen, R. Hilpinen, I. Niiniluoto, and M.B. Provence Hintikka, editors, *Essays in Honour of Jaakko Hintikka*, pages 61–72. Boston: Reidel, 1979.

[Cook, 1971] S.A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the 3rd Annual ACM Symposium on Theory of Computing*, pages 151–158. New York: Association for Computing Machinery, 1971.

[Craig, 1957] William Craig. Three uses of the hebrand-gentzen theorem in relating model theory and proof theory. *Journal of Symbolic Logic*, 22:269–285, 1957.

[Dummett, 1973] Michael Dummett. *Frege: Philosophy of Language*. London: Duckworth, 1973.

[Dunn and Jr, 1968] J.M. Dunn and N.D. Belnap Jr. The substition interpretation of quantifiers. *Nous*, 2:177–185, 1968.

[Fitch, 1952] F.B. Fitch. *Symbolic Logic*. New York: Ronald Press, 1952.

[Frege, 1879] Gottlob Frege. *Begriffsschrift, A Formula Language, Modeled upon that of Arithmetic, for Pure Thought*. Cambridge, MA: Harvard University Press, 1879.

[Gamut, 1991] L.T.F. Gamut. *Logic, Language and Meaning, volume 2 [= Intensional Logic and Logical Grammar]*. Chicago: University of Chicago Press, 1991. L.T. F. Gamut is the collective pseudonym for Johan van Benthem, J.A Groenendijk, D.H. de Jongh and M.J. Stokhoff.

[Garey and Johnson, 1979] M.R. Garey and D.S. Johnson. *Computers and Intractability*. San Fransisco: W.H. Freeman, 1979.

[Gentzen, 1934] Gerhard Gentzen. Untersuchungen uber das logische Schliessen. *Math. Zeitschrift*, 39:176–210 and 405–431, 1934.

[Gödel, 1931] Kurt Gödel. Über formal unentscheidbare Satze der Principia Mathematica und verwandter Systeme I. *Montashefte fur Mathematik und Physik*, 38:173–198, 1931. translated in van Heijenoort *From Frege to Godel: A Source Book in Mathematical Logic 1879–1931* published in Cambridge, MA by Harvard University Press in 1967; pages 596–616.

[Goldfarb, 1979] Warren Goldfarb. Logic in the twenties: The nature of the quantifier. *Journal of Symbolic Logic*, 44:351–368, 1979.

[Hacking, 1979] Ian Hacking. What is logic? *Journal of Philosophy*, LXXVI:285–319, 1979.

[Harman, 1986] Gilbert Harman. *Change in View*. Cambridge, MA: MIT Press, 1986.

[Henkin, 1961] Leon Henkin. Some remarks on infinitely long formulas. In *Infinistic Methods: Proceedings of the Symposium on the Foundations of Mathematics, Warsaw*, pages 167–183. London: Pergamon, 1961.

[Hilbert and Ackermann, 1928] David Hilbert and Wilhelm Ackermann. *Grundzuge der theoretischen Logik*. Berlin: Springer Verlag, 1928.

[Hilbert, 1923] David Hilbert. Die Logischen Grundlagen der Mathematik. *Math. Annalen*, 88:151–165, 1923.

[Hintikka, 1955] Jaakko Hintikka. *Two Papers on Symbolic Logic*. Helsinki: Akateeminen Kirjak, 1955.

[Hintikka, 1987] Jaakko Hintikka. The fallacy of fallacies. *Argumentation*, 1:211–238, 1987.

[Hintikka, 1997] Jaakko Hintikka. What was Aristotle doing in his early logic, anyway? a reply to Woods and Hansen. *Synthese*, 113:241–249, 1997.

[Hodges, 1983] Wilfrid Hodges. Elementary predicate logic. In Dov M. Gabbay and Franz Guenthner, editors, *Handbook of Philosophical Logic: Elements of Classical Logic*, volume 1 of 4, pages 1–131. Dordrecht and Boston: Reidel, 1983.

[Jáskowski, 1934] Stanislav Jáskowski. On the rules of supposition in formal logic. *Studia Logica*, 1:5–32, 1934.

[Kalish and Montague, 1964] Donald Kalish and Richard Montague. *Logic: Techniques of Formal Reasoning*. New York: Harcourt, Brace and World, 1964.

[Kleene, 1952] Stephen Cole Kleene. *Introduction to Metamathematics*. New York: Van Nostrand, 1952.

[Koslow, 1992] Arnold Koslow. *A Structuralist Theory of Logic*. New York: Cambridge University Press, 1992.

[Lear, 1985] Jonathan Lear. *Aristotle and Logical Theory*. Cambridge: Cambridge University Press, 1985.

[Lemmon, 1978] E.J. Lemmon. *Beginning Logic*. Indianapolis: Hackett, 1978.

[Lewis, 1912] C.I. Lewis. Implication and the algebra of logic. *Mind*, 21:522–531, 1912.

[Locke, 1975] John Locke. *An Essay Concerning Human Understanding*. Oxford: Clarendon Press, 1975. Originally published in 1689.

[Massey, 1975] Gerald J. Massey. Are there any good arguments that bad arguments are bad? *Philosophy in Context*, 4:61–77, 1975.

[Massey, 1975a] Gerald J. Massey. In defense of the asymmetry. *Philosophy in Context*, 4:44–45, 1975a. Supplementary Volume.

[Massey, 1981] Gerald J. Massey. The fallacy behind fallacies. *Midwest Studies in Philosophy*, 6:489–500, 1981.

[Mates, 1965] Benson Mates. *Elementary Logic*. Oxford: Oxford University Press, 1965.

[Moore, 1980] G.H. Moore. Beyond first-order logic: The historical interplay between mathematical logic and axiomatic set theory. *History and Philosophy of Logic*, 1:95–137, 1980.

[Normore, 1993] Calvin Normore. The necessity in deduction: Cartesian inference and its medieval background. *Synthese*, 96:437–454, 1993.

[Peirce, 1885] Charles S. Peirce. On the algebra of logic: A contribution to the philosophy of notation. *American Journal of Mathematics*, 7:180–202, 1885.

[Peirce, 1992] C.S. Peirce. *Reasoning and the Logic of Things: The 1898 Cambridge Conferences Lectures by Charles Sanders Peirce*. Cambridge, MA: Harvard University Press, 1992. Kenneth Laine Ketner (ed.).

[Popper, 1992] Karl Popper. *The Logic of Scientific Discovery*. New York: Routledge, 1992. Originally published in 1934.

[Prawitz, 1965] Dag Prawitz. *Natural Deduction: A Proof-Theoretical Study*. Stockholm: Almqvist and Wisksell, 1965.

[Prawitz, 1975] Dag Prawitz. Comments on gentzen-style procedures and the classical notion of truth. In J. Diller and G.H. Müller, editors, *Proof Theory Symposium: Lecture Notes in Mathematics 500*, pages 290–319. Berlin: Springer, 1975.

[Putnam, 1975b] Hilary Putnam. Probability and confirmation. In Hilary Putnam, editor, *Mathematics, Matter and Method*, pages 293–304. Cambridge: Cambridge University Press, 1975b. Originally published in *The Voice of America, Forum Philosophy of Science* volume 10 in 1963.

[Putnam, 1975d] Hilary Putnam. An examination of Grunbaum's philosophy of geometry. In *Mathematics, Matter and Method*, pages 93–129. Cambridge: Cambridge University Press, 1975d. Originally published in *Philosophy of Science: The Delaware Seminar 2* by Interscience Publishers of New York. B. Baumrin editor.

[Quine, 1950] W.V. Quine. *Methods of Logic*. New York: Holt, 1950.

[Quine, 1970] W.V.O. Quine. *Philosophy of Logic*. Englewood Cliffs, NJ: Prentice-Hall, 1970. ‘

[Routley *et al.*, 1982] R. Routley, R. K. Meyer, V. Plumwood and R. Brady. *Relevant logics and their Rivals 1*, Ridgeview Publishing Company, 1982.

[Suppes, 1957] Patrick Suppes. *Introduction to Logic*. New York: Van Nostrad Reinhold, 1957.

[Tennant, 1978] Neil Tennant. *Natural Logic*. Edinburgh: Edinburgh University Press, 1978.

[Tennant, 1987] Neil Tennant. *Anti-Realism and Logic*. Oxford: Clarendon Press, 1987.

[Tennant, 1992] Neil Tennant. *Autologic*. Edinburgh: Edinburgh University Press, 1992.

[Thistlewaite *et al.*, 1988] P.B. Thistlewaite, M.A. McRobbie, and R.K. Meyer. *Automated Theorem-proving in Non-classical Logics*. London: Pitman and New York: John Wiley, 1988.

[Thom, 1981] Paul Thom. *The Syllogism*. Munich: Philosophia Verlag, 1981.

[Thomason, 1970] Richmond H. Thomason. *Symbolic Logic*. New York: Macmillan, 1970.

[Urquhart, 1984] Alasdair Urquhart. The undecidability of entailment and relevant implication. *Journal of Symbolic Logic*, 49:1059–1073, 1984.

[Urquhart, 1990] Alasdair Urquhart. The complexity of decision procedures in relevance logic. In J. Michael Dunn and Anil Gupta, editors, *Truth or Consequences. Essays in Honor of Nuel Belnap*, pages 61–76. Dordrecht and Boston: Kluwer, 1990.

[Urquhart, 1999] Alasdair Urquhart. The complexity of decision procedures in relevance logic ii. *Journal of Symbolic Logic*, 64:1774–1802, 1999.

[van Dalen, 1980] D. van Dalen. *Logic and Structure*. Berlin: Springer, 1980.

[Woods and Hansen, 1997] John Woods and Hans V. Hansen. Hintikka on aristotle's fallacies. *Synthese*, 113:217–239, 1997.

[Woods and Hansen, 2001] John Woods and Hans V. Hansen. The subtleties of Aristotle's non-cause. *Logique et Analyse*, 2001. To appear.

[Woods *et al.*, 2000] John Woods, Andrew Irvine, and Douglas Walton. *Argument: Critical Thinking, Logic and the Fallacies*. Toronto: Prentice-Hall, 2000.

[Woods, 2001] John Woods. *Aristotle's Earlier Logic*. Paris: Hermes Science Publications, 2001.

[Woods, 2002] John Woods. *Paradox and Paraconsistency: Conflict Resolution in the Abstract Sciences*. Cambridge: Cambridge University Press, 2002. To appear.

JOHN WOODS

STANDARD LOGICS AS THEORIES OF ARGUMENT AND INFERENCE: INDUCTION

1 INTRODUCTION

Induction is a central focus for a sprawling research community — certainly for everyone to whose interests this *Handbook* is directed. Among the main participants are mathematicians and statisticians who work in Bayesian statistics, information theory, the theory of measurement and ergodics; engineers and computer scientists who work on artificial intelligence, complexity theory and entropy; experimental and cognitive psychologists interested in the theory of choice, risk-behaviour, problem solving, learning measurement and heuristics; management theorists and economists concerned with public choice theory, game theory and probabilistic decision theory; epistemologists and philosophers of science concerned with theory adjudication, rationality theory and the foundations of probability; logicians interested in inductive logic; and in inchoate and rather tentative ways, argumentation theorists and informal logicians concerned with practical, everyday reasoning and advocacy. (Rosenkrantz [1981].)

The focus of the present chapter is ampliative reasoning and argument. In the ampliative mode the conclusion contains information that goes beyond the information conveyed by the premisses. There are two main forms of ampliative reasoning and argument, *induction* and *abduction*. We shall see in due course that inductive logic has not achieved the theoretical stability that can be claimed for deductive logic. Even so, there are developments within inductive logic that can reasonably be thought of as standard. The same cannot yet be said for abductive logic. Though a huge literature has arisen in recent years, it would be premature to expect to find standardness of a sort that one does find in inductive logic. We shall not therefore offer an account of abductive logic as a standard logic of argument and inference. (We try to make some repair of this omission in Gabbay and Woods [2002].)

In its most basic sense, an induction is a non-deductive argument or inference from a sample to a conclusion which *projects* the sample in some way.

One way in which a reasoner or arguer projects from a sample can be schematized as follows

1. In sample S, the Fs are Gs
2. So, Fs are Gs.

A second way is this:

i) In the sample, Fs are Gs

ii) So, the next-observed F will be a G.

The first example illustrates *generalization* by inductive inference, and the second illustrates *prediction* by the same means.

The position from which most going systems of inductive logic begin is a philosophical position known as *inductivism*. On this view, scientific reasoning (and, by extension, some kinds of everyday reasoning) is correctly done when it generalizes or predicts by inductive inference from observed data. The general presumption is that an inductive inference is strong to the degree that the data confer increased likelihood on the generalizations and predictions drawn from it. A central task of inductive logic therefore is to set out in a principled way the conditions under which such increases in likelihood occur or, equivalently, an inference is inductively strong.

Bayesianism is a philosophical position according to which likelihood in the intuitive sense just noted is *probability* defined by the axioms of the standard mathematical account of that notion, also known as the probability calculus. It is possible to be an inductivist without being a Bayesian. Had Bayesianism been known in his day, it is quite certain that Francis Bacon [1905] would not have been a Bayesian, even though he was a strong inductivist. Similarly it is possible to hold that the intuitive notion of likelihood is properly identified with probability and yet to refuse the idea that inductive reasoning is the heart and soul of properly transacted science. Karl Popper [1992], for example, is a strong anti-inductivist, though sometimes he is a Bayesian about likelihood. However, inductivism and Bayesianism are the jointly dominant approacher to the philosophy of science in the 20th century and beyond, and this has obvious consequences for the logic of that structure which inductivists take to be fundamental to correct scientific reasoning.

In chapter 1 we noted Mill's remark that induction, when properly executed, is too complex a mode of reasoning and arguing for individual agents to manage. We shall return to this question of complexity in due course.

2 INDUCTION AND HUME'S PROBLEM

In the philosophical literature induction is, as we have said, centrally a matter of generalization or prediction from instances of types of phenomena. A generalization, e.g. that all ravens are black, is taken to have a likelihood that increases with the number (and perhaps the mix) of its positive instances, that is, discrete data in the form, "This, that, and the other raven is black". Underwriting any such conviction is the *Principle of Induction*.

In a broader version, the principle also guides our statistical inferences from samples to populations. Given that n% of a properly constituted sample has property Ψ, we conclude that n% (or approximately n%) of the sample's entire population likewise has property Ψ. In the special circumstance in which the population is taken to be the result of just one case added to the sample in question, that this will also show property Ψ is underwritten by the *Principle of Singular Induction*.

What justifies our confidence in these principles? Consider the generalization G, "All ravens are black". What reason is there to think that ravens not yet examined will conform to G? If we say that previously unexamined ravens come out under eventual inspection to conform to it or that other generalizations have won confirmation in this way, we now have the task of justifying the further generalization that generalizations with positive instances and without negative instances so far are well-justified. Taking this for granted comes to saying, question-beggingly, that the Principle of Induction is justified by itself. (Circle 1, let's call it.) So it might be if the Principle were a truth of logic or even self-evident in some way. But such appears not to be so, in as much the proposition, "Any generalization justified by the Principle may be untrue" is logically consistent. On the other hand, if we hold that the Principle, in its application to future or unexamined cases, will hold true because it has probably done so up to now, we will have landed ourselves in a circle, tight enough to be a dot. (Circle 2) This is, in all essentials, Hume's Problem of Induction and Hume's point was that arguments apparently licensed by the Principle of Induction aren't arguments at all, but rather inductive habits, some of which may *strike* us as having more "survival value" than others, but none of which is objectively more reasonable or justified than any other. Thus

> I shall add, for a further confirmation of the foregoing theory, that, as this operation of the mind, by which we infer like effects from like causes, and *vice versa*, is so essential to the subsistence of all human creatures, it is not probable, that it could be trusted to the fallacious deductions of our reason, which is slow in its operations; appears not, in any degree, during the first year of infancy; and at best is, in every age and period of human life, extremely liable to error and mistake. It is more conformable to the ordinary wisdom of nature to secure so necessary an act of the mind, by some instinct or mechanical tendency, which may be infallible in its operations, may discover itself at the first appearance of life and thought, and may be independent of all the laboured deductions of the understanding. (Hume [1902, Section V, Part II, Paragraph 45])

Hume's skepticism bears upon the issues of the present chapter in a telling way. He can be taken as having developed a position on the relationship of

logic to the argument-inference pair, to wit: Inference is instinctual and not a matter of concocting some or other argument-architectonic. In fact, much of human inference may be sublinguistic. Though there may be a logic of arguments, no would-be theory of our inferential habits qualifies as logic. Whereupon no rule of logic is a correct "rule of inference".

Worse still, Hume's Problem is unsettling for empirical knowledge. If induction is the basis of such knowledge, no set of mere facts can attest to its effectiveness, on pain of circularity. This skepticism also cuts in the opposite direction, in the direction, that is to say, of an Anti-Induction Principle by which the more often inductive methods have *failed* in the past the more likely they will work in future (cf. the well-known Gambler's Fallacy). If it is no more reasonable to bet on the Principle of Induction that on that of Anti-Induction, our habitual affection for the former and our only very occasional acquiescence in the latter are shown to be rationally unfounded. This lack of rational mooring for our standard epistemic practices extends well beyond the alleged fact that there can be no logic of such practices. It would also appear that empirical theories of them are likewise blocked. For what would a theory of inductive inference be if not a theory of how we adapt to and learn from experience? Unless the environments to which we adapt and from which we learn are available to us as pure "givens", then that to which we adapt and from which we learn must itself be learned or inferred, and thus involve inescapably inductive description. Inductive methods are essential to the most basic descriptions of the world and thus are unavoidably employed by any theory determined to establish their effectiveness and truth-tropicity (which is Lipton's word for truth-approximation. (Lipton [1991, 8].)) So it would appear that the circularity recurs at this level, too. (Is it circularity 1 or 2?) (Burks [1977].)

If for these reasons there is no hope of a non-circular defense of the legitimacy of inductive inferences, what is there left to say about the legitimacy of inductive *arguments*. It is not enough to hold that since arguments and inferences are different things, arguments escape the heavy weather of Humean skepticism. Either inductive arguments employ the Principle of Induction or they don't. If they do they will feel the teeth of Hume's bite. If they don't, it is open to question whether good inductive arguments, whatever they turn out to be *in fine*, are of any use to human reasoners. For, as in the case of deduction, we have not shown (or said) that since *arguments* are different from *inferences*, deductive logic is of no use to human reasoners. As we saw at the beginning of the preceding chapter there are constraints, including monochronicity, under which *modus ponens* is a good rule, even though it is not so universally. Why could the same not be true of the rules of an inductive logic which managed to produce a satisfactory response to Hume's Problem?

The Principle of Induction in its simplest and most appealing form carries with it the idea of a positive instance of a generalization; as with black

ravens, which are positive instances of generalization G. By an obvious syntactic similarity so too are nonblack nonravens, which are positive instances of the generalization, "All nonblacks are nonravens", which is equivalent to G by contraposition. Why should the confirmation of a generalization depend upon the particular terms in which it chances to be formulated or in which we happen to think it up?

So nonblack nonravens should also and equally qualify as positive instances of G, and so too for black nonravens.[1] Now everything whatever crops up in one or other of the four sets, black ravens, black nonravens, nonblack ravens and nonblack nonravens. From which we have it that everything whatever is either a positive or negative instance of any generalization whatever. Put another way: *conformity is confirmation*; [2] that is, every observation consistent with a generalization confirms it. This, the so-called Paradox of Confirmation, is due to Hempel [1945]. Hempel's own solution is a proposal in effect to swallow the conformity-is-confirmation line on grounds that its counterintuitiveness is a "psychological illusion". (Hempel [1945, section 5.2].) An alternative solution, inspired by Popper [1962] and developed by Watkins [1957] would require us to narrow the idea of confirmation in reaction to Hempel's proposal to broaden it. This is done by restricting the confirmation of an hypothesis to its verified predicted consequences. Hempel's approach provides that any statement implying a second statement which confirms a third statement confirms the third. But Hempel's condition rests uneasily with the Popper-Watkins condition that hypotheses are confirmed by their predicted consequences, for jointly the principles imply that any statement confirms any other.[3] In this we see that the apparent over-broadness of the Hempel approach is not counterbalanced by the apparent over-narrowness of the Popper–Watkins approach, so there is no chance of their cancelling out each others' excesses and leaving us with a shared and more subdued core notion.

Even so, no scientist or everyday reasoner would ever set out to harness his confirmatory practices to Hempel's conditions. This suggests that, narrowness aside, Popper-Watkins is the better approach. That this is so has been challenged, if not outright discredited, by Goodman [1983], where it is argued that the P-W principle cannot discriminate between contrary hypotheses. Let 'grue' be the predicate that applies to green things observed before time t and to blue things not observed before t, where t is some arbitrary selected future date (thus to things blue before t and not observed to be so before t and to things blue after t and not observed to be so before t.)

[1]This is clear when "All ravens are black" is rewritten as "$\forall x(Rx \supset Bx)$", of which "$R \wedge Ba$" is a confirming instance.

[2]In a nice turn of phrase of Rosenkrantz [1981].

[3]*Proof*: Let $\perp$ be a logical truth. Then for all Φ, $\Phi \vdash \perp$. Hence, by P-W, Φ is confirmed (by $\perp$). Let Ψ be any statement distinct from Φ. Then $\Psi \vdash \perp$, and $\perp$ confirms Φ. Hence, by the Hempel condition, Ψ confirms Φ. Every statement confirms every other statement.

A green emerald examined before t is a grue emerald and so is (Hempelian) confirmation of the generalization "All emeralds are grue". Equally confirmed is "All emeralds are green". The two generalizations are incompatible with one another since, after t, they satisfy contrary predictions. What is more, since for Hempel confirmation is *closed under consequence*, examination of green emeralds prior to t confirms the claim that emeralds examined after t will be blue. As for the P-W principle that an hypothesis is confirmed by its consequences, since both "All emeralds are green" and "All emeralds are grue" have the same true consequences at any time prior to t, there is no way to adjudicate between the two conflicting hypotheses at any time prior to some arbitrary time.

Goodman does not regard his problem as having overturned Hempelian confirmation. Goodman's own view is that this New Riddle of Induction calls attention to the necessity of distinguishing lawlike generalizations for which it doesn't hold and shouldn't. The difference turns on entrenchment. Green things are better entrenched grue things and this, thinks Goodman, gives us good reason to generalize to the generalization that all emeralds are green, rather than to the generalizations that they are all grue.

Virtually all commentators take the constructibility of strange predicates such as 'grue' to be the heart and soul of Goodman's problem. In fact, the New Riddle of Induction is a case of Hempel's Paradox. The green emeralds examined before t are positive instances of "All emeralds examined after t are blue", in the same way that yellow lemons are Hempelian instances of "All ravens are black", a generalization that involves no "funny" predicates. How, then, it is possible to draw, without circularity, the distinction between law-like and accidental generalizations in a language all of whose predicates are entrenched in Goodman's sense?

As with Hume's Problem, it is left as a condition on any inductive logic that it make satisfactory provision for the Paradox of Confirmation and its special case, the New Riddle of Induction. (Hempel and Goodman are discussed further in Woods [2001e, chapter 9]. See also Stalker [1994].)

With these things said, a problem arises for the very prospect of a *logic* of induction. If Hume's Problem of Induction and its heirs, such as Goodman's New Riddle, are left without a satisfactory rebuttal, the possibility exists that our inductive practices *cannot* be justified, never mind that they are indispensable and the subject of our considerable confidence in actual reasoning. Logic, on the other hand, is widely conceived of as a normative enterprise in which arguments of certain types and inferences of certain patterns are said to be justified to the extent that they conform to various logical *rules*.[4] Thus if there are successful arguments to the effect that induction cannot be justified, they are also arguments to the effect that there can be no such thing as a logic of induction.

[4]For all its orthodoxy, such a view is queried in Chapter 1 above.

This leaves the would-be theorist with two broad options:

Option One: The theorist can provisionally give up on the idea of a logic of induction as a normative theory and replace it with a modified notion in which a logic of induction is a systematic account of our more or less settled practice considered *descriptively.* (See here Lipton [1991, chapter 1].)

Option Two: The theorist might attempt to make the case that our actual inductive practices are made normatively secure not by their fidelity to principles called into question by arguments such as Hume's but rather by their fidelity to principles that owe their justification to the principles that Hume didn't worry about. On one such attempt, our inductive practices are justified, at least in part, by their conformity to the calculus of probability whose own justification is secured by the fact that it is a (specialized) branch of the arithmetic of the real line. Another possibility is to construe the property of inductive strength in such a way that its theory is a *deductive* theory and thus a theory that satisfactorily answers Hume's contention that induction cannot have a deductively based justification or logic. (See here Carnap [1962]).

In what follows the reader is invited to interpret the theories under review as sanctioned by option one or option two as the case may be. The general format of our discussion will be this: We shall introduce the concept of an *inductive logic in the broad sense* (IL^B) which is represented as a quadruple ⟨Provisional definition of inductive strength, the probability calculus, inductive logic *in the narrow sense* (IL^N), a philosophical reply to Hume's problem⟩. Inductive logic in the *narrow* sense is motivated by the fact that the probability calculus fixes the inductive strength of very few arguments or inferences which we would regard as inductively strong (never mind, for now, whether we are *justified* in so regarding them). Thus an IL^N is a theory designed to supplement the theorems of the probability calculus with rules (or "rules") for the determination of inductive strength.

The first component of IL^B is furnished by the working definition

> **DefIndSt**: An argument $\langle\{\Phi_1,\ldots,\Phi_n\},\Psi\rangle$ is inductively strong if and only if the probability of Ψ given the Φ_i is higher than the probability of Ψ considered alone and $\langle\{\Phi_1,\ldots,\Phi_n\},\Psi\rangle$ is invalid.

The definition invokes the concept of probability, and probability is the subject of the calculus of probability or **ProbCal**, to which we now turn. But note, in passing, the force of Raven and Grue. "What are the constraints on the Φ_i?"

3 PROBABILITY

In the several pages to follow we shall give a formal (i.e., a mathematical) definition of probability. It is important to note that the formal definition underdetermines the conceptual meaning of probability, concerning which there are a number of rival interpretations. In most interpretations probabilities are defined over *probability spaces* whose members are events. The concept of event varies with different interpretations of the probability concept.

Logical probability On this view, probability is a degree of *rational* belief. Given the same evidence, every rational agent will have the same degree of belief in any given proposition. Early work was done by Johnson [1932] and Jeffreys [1939], but Keynes [1921] is given special recognition. Carnap [1962] is a later contribution to the logical approach.

Subjective probability. Originated by Ramsey [1931] and de Finetti [1964], subjective probabilities are taken as degrees of (rational) belief. An agent's degree of belief is measured by his willingness to bet on it. A belief is "coherent" if it is proof against a Dutch book (see below). An agent, A, resists a Dutch book when it is not possible for any co-bettor to select stakes which guarantees that A will lose no matter what.

Frequency probability. As developed by Richard von Mises (von Mises [1928; 1964]), the probability of an event in a given situation is definable via the frequency of occurrences of events of the same kind in an imaginary population of recurrences of the same situation. In its most elementary form, the frequency theory deals with finite frequencies. The probability of event F occurring given that it is B, is the number of As that we observe B divided by the number of Bs. In the so-called "long run" version of the theory, the probability that A will occur given that it is B is the *limit* that the relative frequency of As in Bs would converge on provided there were indefinitely many Bs. It is open to question, however, whether this limit is what reasoning agents actually aim at when they make judgements of probability. It is a problem that provoked Keynes' celebrated wisecrack that in the long run we are all dead.

Chance probability. Here a probability is an objective (i.e., non-doxastic) non-relational property of events. The property is called *chance*. In virtue of its non-relational character, the *objective* theory of objective chances resembles the subjective probability theory, which is also non-relational (or single-case). But objective chance theories also make use of relative frequencies, which it is the role of the property of chance to explain. Thus chance accounts for how relative frequencies are evidentially fruitful, (Mellor [1971] and Lewis [1980]). The theory proposes that chance determines what degree of belief is appropriate without *itself* being a relative frequency or a degree of belief. In this regard, propensity theories are also classified as frequency accounts. On this view, probability is a propensity that inheres

in a group of repeatable events. The probability r of a particular event of kind k is a frequency of k-events close to r given the propensity that inheres of repretitions of the salient repetitions.

Subsuming the logical, subjective and objective theories is an even more basic distinction adumbrated in Hacking's remark that probability is a Janus-faced concept.

> [P]robability ... is Janus-faced. On the one side it is statistical, concerning itself with stochastic laws of chance processes. On the other side, it is epistemiological, dedicated to assessing reasonable degrees of belief in propositions quite devoid of statistical background (Hacking [1975, 12]).

Thus there are two fundamental conceptions of probability, in one of which probabilities are in the head and, in the other, are in the world. This is a loose way of talking even that heads are also in the world and that the in-relation is transitive. There is no fully settled nomenclature for marking the intended distinction; but we could do worse than *material* probability for "in the world" and *epistemological* probability for "in the head". So taken, logical and subjective theories are espistemological, and frequency and objective theories are material.

Other interpretations still have been made of the concept of probability, making it a semantically layered notion and occasion of theoretical rivalry. (See Howson [1995] for an excellent survey.) Even so, three interpretations lightly sketched here and most of the others (an exception is considered below) satisfy the axioms of the probability calculus with little or no adjustment. So we shall turn now to probability theory.[5]

4 THE PROBABILITY CALCULUS

The mathematical theory of probability was pioneered by Kolmogorov [1950].[6] A good account of modern developments can be found in Billingsley [1979].

In deductive logic, an argument $\langle\{\Phi_1, \ldots, \Phi_n\}, \Psi\rangle$ is valid if and only if its corresponding conditional, $\ulcorner(\Phi_1 \wedge, \ldots, \wedge\Phi_n) \supset \Psi\urcorner$, is a necessary truth; that is, if the premisses conjoined entail Ψ. By this test, a considerable number of arguments, including most that have ever been or will be advanced by real people in real-life situations, are bad arguments. Of these, a considerable

[5]For an interesting attempt at reconciling objective and subjective conceptions, see [Atkinson and Peijnenburg, 1999].

[6]The original Kolmogorov axioms are, for any proposition Φ and Ψ:
$0 \leq Pr(\Phi) \leq 1$
If $Pr(\Phi) = 1$, then Φ must be true
If Φ and Ψ are incompatible, then $Pr(\Phi \text{ or } \Psi) = Pr(\Phi) + Pr(\Psi)$ (Kolmogorov [1956]).

number are nevertheless, in some serious way, good. Some are said to be *inductively strong*. At least one large class of inductively strong arguments are those that appear to satisfy **DefIndSt**.

Implicit in **DefIndSt** is a distinction between the probability a sentence has independently, and the probability it has in the light of some other sentence or sentences. We may speak of the former probability as "prior", and of the latter probability as "posterior". These ways of speaking are loose conveniences at best. They allow us partially to characterize inductive strength: an argument $\langle\{\Phi_1, \ldots, \Phi_n\}, \Psi\rangle$ is inductively strong only if Ψ's posterior probability is greater than its prior probability. It must not be thought that, in our example, Ψ's prior probability is the probability it has independently of any given set of conditions. Similarly, Ψ's posterior probability is not, in our example, its value in light of any set of conditions, but only in light of the conditions specified by the argument's premisses; so the *premisses* are an essential factor. In most standard theories of probability, all probabilities are conditional. Thus Ψ's prior probability is its conditional probability independently of the conditional probabilities of the particular premisses Φ_1, ..., Φ_n. Often enough, the conditional nature of prior probabilities is not spelled out. There are reasons for this, some technical and some rather more substantive. On the technical side, suppressing the conditionality of prior probabilities simplifies the formulation of those probabilities in which the theorist has a particular interest. On the substantive side, the condition that fixes the probability of a given sentence which, from the point of view of some further set of conditions, is prior to them, is a condition that is sometimes unknown to the theorist. In some treatments of probability, the theorist postulates a *universe of discourse* **U**, which is represented by a class of disjoint alternatives. Then **U** itself is said to fix the possibilities of each atomic sentence describing the alternatives in **U**. The prior probability, then, of each such sentence is its conditional probability on **U**. Where Φ and Ψ are two such sentences having their prior probabilities fixed by **U**, it is possible to compute the posterior probability of (say) Ψ, *given* Φ. This will be the conditional probability of Ψ on Φ.

Not all posterior probabilities are conditional in this way. For if Φ and Ψ are as before, it is possible to compute the probability of $\ulcorner(\Phi \wedge \Psi)\urcorner$ and of $\ulcorner\Phi \vee \Psi\urcorner$. True, the probability of $\ulcorner\Phi \wedge \Psi\urcorner$ is conditioned by the prior probabilities of Φ and of Ψ, but the sentence to which the conditional probability is assigned in this case is not, for obvious reasons, Ψ(on Φ) or Φ(on Ψ). Similarly, given that Φ has the prior probability that it does have, it is possible to compute the posterior probability of $\ulcorner\neg\Phi\urcorner$.

We have been speaking just now of computing the probabilities of certain statements given the prior probabilities of other statements. This, too, is loose talk. It must not be taken to mean that it is always possible to know the conditional probability of a sentence Ψ on a sentence Φ; for sometimes the prior probabilities of Φ and Ψ won't be known to the theorist. Even so,

where Φ and Ψ have determinate prior values, never mind again whether anyone knows what they are, various conditionalized combinations of these sentences, and of truth functional compounds of them, likewise will have determinate posterior probabilities, never mind whether they are known to anyone. We see, then, that for various kinds of sentences, posterior probabilities are functions of prior probabilities. Those functional relationships are constrained by minimal conditions known as the axioms and theorems of the probability calculus, to a discussion of which we will now turn.

As with various systems of deductive logic, the probability calculus, **ProbCal**, is a couple $\langle G, I\rangle$ of a grammar and a set of interpretations. The grammar gives a vocabulary and a set of Formation Rules. The vocabulary of **ProbCal** is made up of (i) an arbitrarily large number of atomic sentences, p, q, r, s, p_1, …, p_n; (ii) the connectives, $\vee$, $\wedge$ and $\neg$; (iii) the real-valued variables, x, y, z, w_1, x_1, …, x_n, …; the functional expression '*Pr*'; (iv) and the elementary vocabulary of analysis, i.e., the arithmetic of the real line. The formation rules include (i) the sentence-forming rules of **PC** as applied to our connectives; (ii) the descriptor rule: if Φ is a sentence then $\ulcorner Pr(\Phi)\urcorner$ is a denoting phrase which is read "the probability of Φ"; (iii) the formation rules of arithmetic.

Each $i \in I$ is an interpretation of **G**. Given an interpretation, each sentence of **ProbCal** has exactly one truth-value, T or F. The variables x, y, z, x_1, x_2, etc., range over the real numbers in the open interval 0 to 1. Each sentence of **ProbCal** has a unique assignment of a real number in $[0, \ldots, 1]$ and the denoting phrase "$Pr(\Phi)$" denotes that number. The standard interpretation of the elementary syntax of the arithmetic of the real line is also assumed. As here presented, **ProbCal** is a variation of the treatment of Burks [1977] which, in turn, derives from Hosiasson-Lindenbaum [1940]. A further, and accessibly "Burksian" version, is Gustason [1994].

Before turning to the probability axioms of **ProbCal**, it is necessary to introduce the concept of *mutual exclusivity*.

> **MutExc**: If Φ_1, …, Φ_n and Ψ are **ProbCal** sentences, then Φ_1, …, Φ_n are mutually exclusive on condition Ψ if and only if at most, one of the Φ_i can be true if Ψ is true.

Some sentences, χ_1, …, χ_i, are mutually exclusive no matter what, on any condition. And sometimes the condition on which a set of sentences is mutually exclusive is obvious and virtually guaranteed to obtain. In such cases, the rider "on condition so-and-so" may be omitted. We may now state our first axiom.

AXIOM 1. If Φ_1, …, Φ_n are mutually exclusive given condition Ψ then $Pr(\Phi_1 \vee \ldots \vee \Phi_n/\Psi) = \sum_{i=1}^{n} Pr(\Phi_i/\Psi)$.
(That is, the probability of a disjunction on a condition is the sum of the probabilities of each disjunct on that same condition.)

Suppose we were to roll a standard six-sided die. Our disjunctive proposition might be: "The result of the roll is a one *or* a three *or* a five". Each disjunct (i.e. each individual proposed result) has its own probability given the background conditions Ψ (i.e. that we roll a six-sided die, each of whose faces has a given chance of being the result of our roll). The probability of the first disjunct is primitively assigned as $1/6$, as are the following two. Thus $Pr(\Phi_1/\Psi) = 1/6$, $Pr(\Phi_2/\Psi) = 1/6$, and $Pr(\Phi_3/\Psi) = 1/6$. The sum of these probabilities is thus $1/6 + 1/6 + 1/6 = 3/6 = 1/2$. So the probability of "the result of the roll is a one or a three or a five" is$1/2$.

The second axiom tells a similar story about the conditional probability of conjunctions.

AXIOM 2. $Pr(\Phi_1 \wedge \ldots \wedge \Phi_n/\Psi) = Pr(\Phi_1/\Psi) * \ldots * Pr(\Phi_n/\Phi_1 \wedge \ldots \Phi_n - 1 \wedge \Psi)$.

(In this exposition, multiplication is denoted by the asterisk $*$.)
This means that the probability of a conjunctive proposition being true is the product of the probability of the first conjunct given the background conditions, Ψ, and the second conjunct given both the occurrence of the first and the background conditions, and the probability of the third conjunct given the occurrence of the first two and the background conditions, and so on.

Suppose we have an ordinary deck of 52 cards. We may use this axiom to calculate the probability of drawing three jacks in a row. Our conjunctive proposition is thus "drawing a jack *and* a jack *and* a jack from a deck of 52 cards without replacement". The probability of Φ_1 given Ψ is $4/52$, since any one of the four jacks pulled from the 52 cards will satisfy this condition. The probability of Φ_2 given the occurrence of Φ_1, and of Ψ, will be $3/51$, since any one of the three remaining jacks pulled from the 51 remaining cards will satisfy this conjunct. The probability of Φ_3 given the occurrence of Φ_1 and Φ_2 and of Ψ is $2/50$, since either of the two remaining jacks pulled from the 50 remaining cards will satisfy this condition. The probability of $(\Phi_1 \wedge \Phi_2 \wedge \Phi_3)$ is thus $4/52 * 3/51 * 2/50 = 24/132{,}600 = 1/5525$.

Axioms 1 and 2 are not schematic statements of inductive probability. A statement of *inductive probability* is a statement in the form

> Evidence E supports proposition P

or

> Given evidence E, it is probable that P.

A statement of the *probability calculus* is a statement in the form

> The proposition "Evidence E supports proposition P to degree x" is true.

This is not *quite* right even though it captures the right idea. In the way we have explained it just now, a statement of the probability calculus is

a metalinguistic statement, a statement about the statement flanked by quotation marks. This is not strictly speaking true. In all standard formulations '*Pr*' is a term-forming operator on sentences, not an expression of the metalanguage. Thus a statement in the form $Pr\ (\Phi/\Psi) = x$ does not name the statements in (Φ/Ψ). Even so, that statement is true just in case the conditional probability of the statement named by Φ on the statement named by Ψ is in fact x.

Axiom 3 establishes that deductive entailment is a limiting case of probability. Putting '$\models$' for 'entails',

AXIOM 3. If $\Phi \models \Psi$ then $Pr\ (\Psi/\Phi) = 1$, *except* where Φ is a logical falsehood.[7]

The exception is explained as follows. The probability calculus assumes a standard logic for deduction. Accordingly, **ProbCal** countenances the classical metatheorem known as *ex falso (sequitur) quodlibet*: a contradiction entails any proposition whatever. If in Axiom 3 Φ were allowed to be a contradiction, then every statement whatever would have a probability of 1 relative to Φ. In particular, we would have $Pr(\Psi/\Phi) = Pr(\neg\Psi/\Phi) = 1$. But by Axiom 1, $Pr(\Psi \vee \neg\Psi/\Phi) = 1 + 1 = 2$, which violates the condition that there is no probability value greater than 1.

Axiom 4 asserts that probability values on a condition Φ remain the same for all conditions logically equivalent to Φ.

AXIOM 4. If $\Phi \models \Psi$ and $\Psi \models \Phi$ then for all χ,

$$Pr(\chi/\Phi) = Pr(\chi/\Psi).$$

The fifth and last axiom merely stipulates that there are no *negative* probability values. That is

AXIOM 5. $Pr(\Psi/\Phi) \geq 0$.

We turn now to some theorems of **ProbCal**.

THEOREM 1. *If Φ is a logical truth then for any Ψ,* Pr *$(\Phi/\Psi) = 1$.*

Proof. Assume that Φ is a logical truth. (Then $\Psi \models \Phi$ by deductive logic.) Hence, by Axiom 3, $Pr(\Phi/\Psi) = 1$, unless Ψ is a contradiction. ■

Theorem two introduces the connective $\neg$.

THEOREM 2. Pr*$(\neg\Phi/\Psi) = 1$*-Pr*(Φ/Ψ)*

Proof. It is a truth of logic that $\ulcorner(\Phi \vee \neg\Phi)\urcorner$. Thus, by Theorem 1, $Pr(\Phi \vee \neg\Phi/\Psi) = 1$. It is a truth of deductive logic that $\Psi \models \neg\ulcorner(\Phi \wedge \neg\Phi)\urcorner$. Hence, by definition, Φ and $\ulcorner\neg\Phi\urcorner$ again, are mutually exclusive on Ψ. By axiom 1, $Pr(\Phi \vee \neg\Phi/\Psi) = Pr(\Phi/\Psi) + Pr(\neg\Phi/\Psi)$. Given ordinary arithmetic

[7]In which case $Pr(\Psi/\Phi)$ is undefined.

with respect to our previous lines, we have it that $Pr(\neg\Phi/\Psi)$=1-$Pr(\Phi/\Psi)$. (Assuming throughout that Ψ is not a contradiction. For ease of exposition, we retain this assumption, except where the contrary is noted.) ■

This theorem allows us to calculate the probability of some Φ's not obtaining under condition Ψ if we know the probability of Φ's obtaining under condition Ψ. This we do by subtracting this latter probability from 1. For example, let Φ represent "pulls a jack from the deck" and Ψ represent "the standard deck of 52 playing cards". As we know, the probability of Φ/Ψ is 4/52, as there are four jacks in 52 cards. The probability of *not* pulling a jack from the deck of 52 cards, (i.e. $Pr(\neg\Phi/\Psi)$, is thus 1- 4/52 or 48/52 or 12/13). To take another example, let Φ represent "the coin comes head up" and Ψ represent "a random toss of a standard coin". As we know, the probability of Φ/Ψ is 1/2. The probability of the coin not coming up heads, but rather tails (i.e. $Pr(\neg\Phi/\Psi)$) is thus 1- 1/2 or 1/2.

Theorem 2 is a fruitful derivation. It allows us to prove that logically false statements have the probability value of 0. It also enables us to show that all probability values fall in the open interval | 0-1 |. Thus, where $\models\Phi$ means that Φ is a logical truth,

THEOREM 3. *If* $\models \ulcorner\neg\Phi\urcorner$ *then for all* Ψ, $\mathrm{Pr}(\Phi/\Psi)=0$.

Proof. Assume that $\models\neg\Phi$. Then, by Theorem 1, $Pr(\neg\Phi/\Psi) = 1$. From which, by Theorem 2 and ordinary arithmetic, $Pr(\Phi/\Psi)$ = 1-1= 0. ■

THEOREM 4. $0\leq \mathrm{Pr}(\Phi/\Psi)\leq 1$.

The proof of Theorem 4 comes readily from Axiom 5 and Theorem 2.

Theorem 4 allows us to preserve Φ's conditional probability on Ψ for any further condition χ equivalent to Ψ. Our next theorem asserts that the conditional probability of Φ on some condition χ is preserved on that condition for any proposition logically equivalent to Φ. Thus

THEOREM 5. *For all* Φ *and* Ψ, *if* $\Phi\models\Psi$ *and* $\Psi\models\Phi$, *then* $\mathrm{Pr}(\Phi/\chi)= \mathrm{Pr}(\Psi/\chi)$.

Proof. Assume that $\Phi\models\Psi$ and $\Psi\models\Phi$. Then by deductive logic, $\ulcorner(\Phi\equiv\Psi)\urcorner \models \ulcorner(\chi\supset(\Phi\vee\neg\Psi))\urcorner$. In fact, by deductive logic, $\chi\models\ulcorner(\Phi\vee\neg\Psi)\urcorner$. By Axiom 3, $Pr(\Phi\vee\neg\Psi/\chi)$=1. By deductive logic again, $\ulcorner(\Phi\equiv\Psi)\urcorner\models(\chi\supset\neg(\Phi\wedge\neg\Psi))$, hence, given that $\ulcorner(\Phi\equiv\Psi)\urcorner$ is a logical truth, $\chi\models\ulcorner\neg(\Phi\wedge\neg\Psi)\urcorner$. Thus, Φ and $\ulcorner\neg\Psi\urcorner$ are mutually exclusive on χ. Given **Ax1**, $Pr(\Phi\vee\neg\Psi/\chi) = Pr(\Psi/\chi) + Pr(\neg\Psi/\chi)$. By ordinary arithmetic, we obtain $Pr(\Phi/\chi)$=1-$Pr(\neg\Psi/\chi)$. By Theorem 2 and arithmetic, $Pr(\Phi/\chi)$ = 1-$Pr(\neg\Psi/\chi)$. Whereupon, by the transitivity of identity, $Pr\ (\Phi/\chi)$=$Pr(\Psi/\chi)$. ■

We note that Axiom 1 holds only for mutually exclusive alternatives. Of course, not all possibilities *are* mutually exclusive, and so we need a theorem that recognizes this fact.

THEOREM 6. Pr*($\Phi_1 \vee \ldots \vee \Phi_n/\Psi$)=1*-Pr*($\neg\Phi_1 \wedge \ldots \wedge \neg\Phi_n/\Psi$).*

Proof. By Theorem 2 and ordinary arithmetic $Pr(\Phi_1 \vee \ldots \vee \Phi_n/\Psi) =$ 1-$Pr(\neg(\Phi_1 \vee \ldots \vee \Phi_n)/\Psi)$. But $\ulcorner\neg(\Phi_1 \vee \ldots \vee \Phi_n)\urcorner$ is logically equivalent to $\ulcorner(\neg\Phi_1 \wedge \ldots \wedge \neg\Phi_n)\urcorner$. Hence, by Theorem 5 $Pr(\Phi_1 \vee \ldots \vee \Phi_n/\Psi) =$ 1-$Pr(\neg\Phi_1 \wedge \ldots \wedge \neg\Phi_n/\Psi)$. ■

Before attempting to secure this theorem, we should ensure that we clearly understand Axioms 1 and 2, and Theorem 2. The value of the probability of a disjunctive proposition is equal to 1 (the upper limit of probability) *minus* the probability of the conjunction of the negations of each of those propositions which originally occurred as disjuncts. Here is an example.

Suppose we want to know the probability of a 1, 2 or a 3 coming up on the single roll of a standard six-sided die. While we could calculate this directly via Axiom 1. Theorem 6 gives us an alternative method. We first calculate the probability of the conjunction of the negation of the original disjuncts. For instance, suppose Φ_1, Φ_2 and Φ_3 respectively represent the propositions that the roll of the die produces a 1, a 2, and a 3. The probability of a 1 occurring (i.e., $Pr(\Phi_1/\Psi)$) is 1/6. The probability of the negation of this can be calculated via Theorem 2 as being 5/6 (this ought to be somewhat intuitive). Given a fair roll, the chances of obtaining a number which is not 1 is 5/6. $Pr(\neg\Phi_1 \wedge \neg\Phi_2 \wedge \neg\Phi_3/\Psi)$ can be calculated via Axiom 2 to be $Pr(\neg\Phi_1/\Psi) * Pr(\neg\Phi_2/\neg\Phi_1 \wedge \Psi) * Pr(\neg\Phi_3/\neg\Phi_1 \wedge \neg\Phi_2 \wedge \Psi)$, which is $5/6 * 4/5 * 3/4$ or 60/120 or 1/2. Now we may take this result and subtract it from 1. 1 - 1/2 = 1/2. Thus the probability of either a 1, 2 or a 3 coming up on a single roll of a standard six-sided die is 1/2.

We now state a handy *heuristic rule*: If the statement whose probability is to be calculated is a disjunction, are the disjuncts mutually exclusive? If they are, apply Axiom 1; if not, employ Theorem 6.

In the present example, Φ_1 and Φ_2 and Φ_3 are *independent* on the condition Ψ (i.e., that there will be successive tosses of a true coin). Independence is difficult to characterize in a non-technical way. It will serve our purposes well enough to say that $\Phi_1 \ldots, \Phi_n$ constitute a set of independent statements on a condition Ψ just in case in relation to Ψ the probability of no Φ_i is affected by the truth-values of any of the other Φ_j or by any truth-functional compound of pairs of the Φ_j.

Up to now, we have been thinking of probability as a function on (pairs of) sentences. It is convenient also to speak of the probabilities of *events*, that is, of the events that would obtain if those sentences were true.

THEOREM 7. *Let $\Phi_1, \ldots, \Phi_n$ be independent on condition Ψ. Then* $\mathrm{Pr}(\Phi_1 \wedge \ldots \wedge \Phi_n/\Psi) = \prod_{i=1} < n\mathrm{Pr}\phi_i/\psi)$.

Proof. Suppose $\Phi_1, \ldots, \Phi_n$ to be independent on Ψ. By the meaning of independence, $Pr(\Phi_i/\Psi) = Pr(\Phi_i/\Phi_1 \wedge \ldots \wedge \Phi_{i-1} \wedge \Psi)$, for each i such that $2 \leq i \leq n$. Then, by **Ax2** and arithmetic $Pr(\Phi_1 \wedge \ldots \wedge \Phi_n/\Psi) = \prod_{i=1} < nPr\phi_i/\psi)$ ■

This means that when the probability of each proposition in a conjunction is unaffected by whether or not any of the propositions in the conjunction actually obtains (i.e., they are independent of one another), then the probability of the overall conjunction is equal to the product of the probabilities of each individual conjunct.

Suppose, for example, that we wish to know the probability of having a coin come up heads on 100 successive tosses. This could be presented by $Pr(\mathrm{h}_1 \wedge \ldots \wedge h_{100}/s)$ where h_i represents "comes up heads on flip #i" and s represents the standard condition of a fairly flipped coin. Each h_i has the probability of 1/2, and no toss is affected by any other; that is, regardless of how many times heads has come up successfully prior to a given trial, the probability of heads being the outcome of a given trial is still 1/2.[8] Then $Pr(h_1 \wedge \ldots \wedge h_{100}/s) = Pr(h_1/s) * \ldots * Pr(\mathrm{h}_{100}/s)$ or $1/2 * 1/2 * 1/2\ldots$ until we have 100 multiplicands. Incidently, this yields $1/1.267 * 10^{30}$ or $7.8886 * 10^{-31}$, a number indeed very close to 0.

We may now state a second heuristic rule: If the statement whose probability is to be calculated is a conjunction, are the conjuncts independent or not? If they are independent, use Theorem 7; if they are not, use Axiom 2.

Let us say that a set of sentences $\Phi_1, \ldots, \Phi_n$ is *jointly exhaustive* on condition Ψ just in case at least one of them is true if Ψ is, if and only if, that is, Ψ entails $(\Phi_1 \vee \ldots \vee \Phi_n)$.

THEOREM 8. *If $\Phi_1, \ldots, \Phi_n$ are both mutually exclusive and jointly exhaustive on condition Ψ, then $\sum_{i=n}^{n} \mathrm{Pr}(\Phi_i/\Psi) = 1$*

Proof. Assume the Φ_i to be mutually exclusive and jointly exhaustive on Ψ. Then, by the definition of joint exhaustiveness, $\Psi \models \ulcorner(\Phi_1 \vee \ldots \vee \Phi_n)\urcorner$, and by **Ax3**, $Pr(\Phi_1 \vee \ldots \vee \Phi_n/\Psi) = 1$. Given our original assumption, it follows by definition and **Ax1** that
$Pr(\Phi_1 \vee \ldots \vee \Phi_n/\Psi) = \sum_{i=1}^{n} Pr(\Phi_i/\Psi)$.

By arithmetic on this step and the one which immediately precedes it,

$$\sum_{i=1}^{n} Pr(\Phi_i/\Psi) = 1.$$

■

Our theorem provides that if a group of propositions (a) have no overlap and (b) exhaust all alternatives, then the sum of these probabilities is one.

[8]Many people think otherwise, of course; namely that the probability of yet another head, decreases. This is known as the 'Gambler's Fallacy'.

For example, suppose we wish to calculate the probability of drawing any given playing card, say the 7 of diamonds from the standard deck of 52 cards. This probability will be 1/52. We may also state the probabilities of each card being drawn, systematically The probability of drawing the ace of spades is 1/52, that of drawing the 2 of spades is 1/52 and so on until we have exhausted all 52 possible results, each of which has a probability of 1/52. The sum of these probabilities will be 52/52 or 1.

Let us suppose that $\Phi_1, \ldots, \Phi_n$ are mutually exclusive on condition Ψ. Then the statements $\Phi_1, \ldots, \Phi_n$ are *equiprobable* on Ψ if and only if, for each Φ_i and Φ_j, $Pr(\Phi_i/\Psi) = Pr(Phi_j/Psi)$. This takes us to

THEOREM 9. *If* $\Phi_1, \ldots, \Phi_n$ *are mutually exclusive, jointly exhaustive and equiprobable on* Ψ *then for* k = *1, 2,* …, n, Pr$(\Phi_1 \vee \ldots \vee \Phi_k/\Psi)$ = k/n.

Proof. Assume the antecedents of the theorem. Then, by definition and Theorem 8,

$$Pr(\Phi_1 \vee \ldots \Phi_n/\Psi) = \sum_{i=1}^{n} Pr(\Phi_i/\Psi) = 1$$

Given this step and our assumption, it follows by the definition of equiprobability that each $Pr(\Phi_i/\Psi) = 1/n$. By arithmetic and Axiom 1, for $k = 1, 2, \ldots, n$, $Pr(\Phi_1 \vee \ldots \vee \Phi_k/\Psi) = k(1/n) = k/n$. ■

We return to our coin, and stipulate three tosses. If we ask for the probability of the coin showing tails on the first two tosses and heads on the third, Theorem 7 will tell us, since the tosses are independent. (It is $1/2^3$ or 1/8.) Suppose we ask a different question, viz., what is the probability of getting two tails and one head in any triple of tosses, irrespective of the order in which heads and tails actually appear? In the example at hand, it is easy to see that there are exactly three combinations of outcomes which show two tails and a head. Schematically, they are $T_1 \wedge T_2 \wedge H_3$, $T_1 \wedge H_2 \wedge T_3$, and $H_1 \wedge T_2 \wedge T_3$. So our question is really this: What is the probability of occurrence of one of these combinations? The three combinations are mutually exclusive; so we apply Axiom 1 to obtain:

$$Pr((T_1 \wedge \mathrm{T}_2 \wedge H_3) \vee (T_1 \wedge H_2 \wedge T_3) \vee (H_1 \wedge T_2 \wedge T_3)) = (3(1/8)) = 3/8.$$

In less simple cases, determining the number of combinations is correspondingly less obvious. Suppose that we wish to discover the probability that five draws, with replacement (viz. by Theorem 7), from a standard properly shuffled deck will give three clubs and two non-clubs. The number of distinct ways of obtaining exactly 3 clubs in five draws is

$$\binom{5}{3}.$$

The quantity

$$\binom{5}{3}$$

obeys the algebraic equation:

$$\binom{x}{y} = \frac{x!}{y!(x-y)!}$$

where 'n!' is 'n factorial', that is, $n*$(n-1)$*$(n-2)$* \ldots * 1$. Thus for

$$\binom{5}{3}$$

we have

$$\binom{5}{3} = \frac{5!}{3! * 2!} = \frac{5*4*\cancel{3}*\cancel{2}*\cancel{1}}{3*\cancel{2}*1*2*1} = \frac{20}{2} = 10.$$

There are exactly ten different ways of pulling exactly three clubs in any draw of five. The probability of any one of these combinations is

$$\frac{9}{1024}:$$

$$\frac{13}{15} * \frac{13}{52} * \frac{13}{52} * \frac{39}{52} * \frac{39}{52} = \frac{9}{1024}.$$

If we opt for "without replacement", there is a different result, namely

$$\frac{13}{52} * \frac{12}{51} * \frac{11}{50} * \frac{39}{49} * \frac{38}{48} = 900815.$$

The combinations taken together form a mutually exclusive set of statements. Then, with "Comb$_i$" abbreviating "The ith combination occurs", Axiom 1 gives

Pr(any 3-club-2-nonclub) $= Pr(\text{Comb}_1 \vee \ldots \vee \text{Comb}_1 0)$
$= Pr(\text{Comb}_1) + Pr(\text{Comb}_2) + \ldots + Pr(\text{Comb}_1 0)$
$= 10\ (9/1024) = 45/512 = 0.088.$

We conclude the present section with two central theorems about combinations. Let $\Phi_1, \ldots, \Phi_n$ be independent and equiprobable on condition χ. Let $\Psi_1, \ldots, \Psi_n$ likewise be independent and equiprobable on χ. Suppose further that each pair $\langle \Phi_i, \Psi_i \rangle$' is mutually exclusive and jointly exhaustive on χ. Let Comb$_k$ be one of the

$$\binom{n}{m}.$$

combinations for which $m =$ the number of Φ_i's that are true. The combinations here are conjunctions $\chi_1 \wedge \chi_2 \wedge \ldots \wedge \chi_n$ such that for each $1 \leq i \leq n$, $\chi_i = \Phi_i$, or $\chi_i = \Psi_i$. Finally, put it that $Pr(\Phi_i/\chi) =$ z. Then,

THEOREM 10. $\Pr(\mathit{Comb}_k/\chi) = \Pr(\Phi_i/\chi)^m * \Pr(\Psi_i/\chi)^{n-m} = \mathrm{z}^m (1\text{-z})^{n-m}$.

Let

(a) $\Phi_1, \ldots, \Phi_n$ be independent and equiprobable on condition χ.

(b) $\Psi_1, \ldots, \Psi_n$ be independent and equiprobable on condition χ

(c) Each $\langle \Phi_i, \Psi_i \rangle$ is mutually exclusive and jointly exhaustive of χ.

(d) Comb_k be one of the

$$\binom{n}{m}$$

combinations for which $m =$ the number of situations, i, in which a Φ_i is true; $Pr(\Phi_i/\chi) = z$

Then,

$$Pr(\text{Comb}_k/\chi) = Pr(\Phi_i/\chi)^m * Pr(\Psi_i/\chi)^{n-m} = \text{z}^m(1\text{-}z)^n - m.$$

Our theorem says that we may make a shortcut in the calculation of probabilities of combinations of 'events', given a 'target property' Φ, and its negation Ψ, which jointly exhaust the possibilities of a result. The probability of our target property's obtaining is multiplied by itself the same number of times that the property must obtain in our combination, and the probability of the non-occurrence of that property is multiplied by itself the same number of times that our target property should not occur in our combination. These two products are then finally multiplied.

Note that since Φ_i, Ψ_i are mutually exclusive and jointly exhaustive of possible results, $\langle \Phi_i, \Psi_i \rangle$ is logically equivalent to $\langle \Phi_i, \ulcorner \neg \Phi_i \urcorner \rangle$. For example $\langle$heads, tails$\rangle$ in a coin toss is logically equivalent to $\langle$heads, not-heads$\rangle$.

The second part of our short cut, our shorter cut so to speak, uses this negation to invoke Theorem 2 to obtain $z^m(1\text{-}z)^{n-m}$. Here an example is crucial. Suppose we wish to know the probability of obtaining a 5 or a 6 on exactly three of five successive tosses of a standard six-sided die. There are many ways that this may be done, but we may reduce these by looking for a target property such as 'yielding a 5 *or* a 6', which we'll call Φ, and a complementary property of 'yielding' a 1 *or* a 2 *or* a 3 *or* a 4 which will be referred to either as Ψ, or $\ulcorner \neg \Phi \urcorner$, since Ψ and $\ulcorner \neg \Phi \urcorner$ are logically equivalent. Using Axiom 1 we can see that the probability of Φ is $1/6 + 1/6$ or $1/3$ while $Pr(\Psi)$ is $1/6 + 1/6 + 1/6 + 1/6$ or $2/3$, for any one given die-toss. Having reduced our attention to the jointly exhaustive and mutually exclusive occurrences of Φ or Ψ we may use the factorial method, discussed in our example following Theorem 9 to see that there are only ten combinations of Φ and Ψ which could obtain for our five successive die rolls. As such, we may say that the probability of obtaining exactly three rolls of five or six is equal to $Pr(\text{Comb}_1 \vee \ldots \vee \text{Comb}_1 0)$. The first question is then

"What is the probability of any one of these combinations occurring?" Once that is known we may apply Axiom 1 to get the final answer. Theorem 10 allows the calculation to get the final answer. Theorem 10 allows the calculation of the probability of the occurrence of any combination.

First, let us inspect the long route. One such combination might be: $\ulcorner\Phi_1 \wedge \Phi_2 \wedge \Phi_3 \wedge \Psi_4 \wedge \Psi_5\urcorner$. For this combination we need to calculate $Pr(\Phi_1)$[9]$*Pr(\Phi_2)*Pr(\Phi_3)*Pr(\Psi_4)*Pr(\Psi_5)$, which yields $1/3 * 1/3 * 1/3 * 2/3 * 2/3$ or $4/243$. Another such combination is: $\ulcorner\Psi_1 \wedge \Psi_2 \wedge \Phi_3 \wedge \Phi_4 \wedge \Phi_5\urcorner$. So far the combination we need $Pr(\Psi_1)*Pr(\Psi_2)*Pr(\Phi_3)*Pr(\Phi_4)*Pr(\Phi_5)$, which yields $2/3 * 2/3 * 1/3 * 1/3 * 1/3$ or $4/243$. In short, any combination yields a conjunction of probabilities whose overall probability is the produce of its individual probabilities *à la* Theorem 7

Now the shortcut. We might notice in the above example that the principle of commutativity worked in our favour. That is, regardless of the ordering, we multiplied the probability of Φ (i.e. $1/3$) together three times, we multiplied the probability of Ψ (i.e. $2/3$) together twice, and we multiplied these results by one another to get our final result. In all possible combinations of interest to us, this is in fact what we will do. Thus $Pr(\text{Comb}_i) = Pr(\Phi/\chi^3)*Pr(\Psi/\chi^2) = (1/3)^3*(2/3)^2 = 1/27 * 4/9$ - $4/243$. To generalize further, we multiply $Pr(\Phi/\chi)$ by itself the number of times its occurrence is required of all fruitful combinations, we multiply $Pr(\Psi/\chi)$ by itself the number of times its occurrence is required, and then multiply these two products together. The final touch is to note that the number of times that Ψ is required to occur is equal to the number of 'slots' available in a combination (here 5) *minus* the number of occurrences of Φ that are required. Thus we have

$$Pr(\text{Comb}_i) = Pr(\Phi/\chi)^m*Pr(\Psi/\chi)^{n-m}$$

The 'shorter-cut' requires us to take note of the fact that Ψ is logically equivalent to $\neg\Phi$. In our example, for instance, "yields a 1 or a 2 or a 3 or a 4", given the geometry of dice, is logically equivalent to "does *not* yield a 5 and does *not* yield a 6". So $Pr(\text{Comb}_i)$ is also equal to $Pr(\Phi/\chi)^m*$ $Pr(\neg\Phi/\chi)^{n-m}$. By Theorem 2, $Pr(\neg\Phi/\chi) = 1\text{-}Pr(\Phi/\chi)$. So we may substitute $\ulcorner 1\text{-}Pr\Phi/\chi\urcorner$ for $\ulcorner Pr(\neg\Phi/\chi)\urcorner$. This yields

$$Pr(\text{Comb}_i) = Pr(\Phi/\chi)^m*(1\text{-}Pr(\Phi/\chi)^{n-m}$$

Now, if we replace $\ulcorner Pr(\Phi/\chi)\urcorner$ with z, we obtain the formula:

$$Pr(\text{Comb}_i) = z^m*(1\text{-}z)^n - m$$

which is a much smaller mouthful than the fully extended formula. (Big explanation for a tiny, convenient result.)

[9]Actually, $Pr(\Phi_1/\chi)$ but the background statement χ has been suppressed for the most part.

THEOREM 11. *Given the five antecedent conditions as imposed on Theorem 10,*

$$\begin{aligned}&Pr(Comb_1 \vee \ldots \vee Comb\left(\begin{array}{c} n \\ m \end{array}\right)/\chi)\\ &= \left(\begin{array}{c} n \\ m \end{array}\right) * Pr(\Phi_i/\chi)^m * Pr(\Psi_i/\chi)^{n-m}\\ &= \left(\begin{array}{c} n \\ m \end{array}\right) z^m(1-z)^{n-m}.\end{aligned}$$

This theorem is a minor 'tweaking' on Theorem 10. We know by Axiom 1 that when we have a large disjunction-sentence we may calculate its probability by adding the probabilities of its disjuncts. We also know from the antecedent conditions and Theorem 10 that each combination which serves as a disjunct has the same probability value. We also know from arithmetic that when we add the same number to itself, say ten times, we will get the same answer as we do if we multiply that same number by 10. More generally

$$(x_1 + x_2 + \ldots + x_n) = n(x).$$

That is, five 7's added together yield $5 * 7$.

We know from the factorial method that we are looking to find the probability of $\left(\begin{array}{c} n \\ m \end{array}\right)$ combinations. So the probability of one of the combinations actually obtaining will be the probability of any one of these equiprobable combinations added together $\left(\begin{array}{c} n \\ m \end{array}\right)$ times. Hence

$$Pr(Comb_1) \vee \ldots \vee Comb\left(\begin{array}{c} n \\ m \end{array}\right)\chi) = \left(\begin{array}{c} n \\ m \end{array}\right)Pr(Comb_i/\chi)$$

Theorem 10 gives us a formula for calculating $Pr(\mathrm{Comb}_1/\chi)$, so

$$\begin{aligned}&Pr(Comb_1) \vee \ldots \vee Comb\left(\begin{array}{c} n \\ m \end{array}\right)\chi\\ &= \left(\begin{array}{c} n \\ m \end{array}\right) * Pr(\Phi_i/\chi)^m * Pr(\Psi_i/\chi)^{n-m}\\ &= x^m(1-z)^{n-m}\end{aligned}$$

At this juncture readers may be impatient to know what use — apart from some anticipated returns at the tables at Monte Carlo — are the theorems of **ProbCal**. We have already mentioned that the theorems of **ProbCal** do not chart the whole course of correct inductive inference for the would-be thinkers. A person could know the theorems of **ProbCal** like the back of

his hand and still be a decrepit inductive reasoner. Except for the limiting cases of logical truths and logical falsehoods, the theorems of **ProbCal** will not enable the inductive reasoner to attribute a probability value to any statement of conditional probability in which he might be interested. Unless he has knowledge of the relevant prior probabilities, the theorems of **ProbCal** give the reasoner no determinate counsel about what might be inferred from what and with what degree of reliability.

Even so, often the business of practical reasoning (and scientific reasoning, too) involves the adjudication of rival claims or hypotheses, and concerning such adjudications two theorems of **ProbCal** turn out to be directly pertinent. Our theorems will require a preliminary result, whose proof resembles the proof of Theorem 5. Hence

LEMMA 12. *If* $\chi \models (\Phi \equiv \Psi)$ *then* $\Pr(\Phi/\chi) = \Pr(\Psi/\chi)$

Suppose now that $\Phi_1, \ldots, \Phi_n$ are mutually exclusive and jointly exhaustive on χ. For example, they could be the total set of rival forensic claims about who murdered Jones or the total set of competing scientific hypothesis about some class of phenomena. Let χ be a statement expressing the results of a test of these hypotheses. Recall that the hypotheses are mutually exclusive. So, too, are the conjunctions $\Phi_1 \wedge \chi, \ldots, \Phi_n \wedge \chi$, each of which express possible conditions in which χ is true. This leads us to the Prediction Theorem:

THEOREM 13. *If* $\Phi_1, \ldots, \Phi_n$ *are mutually exclusive and jointly exhaustive on* Ψ*, then*

$$Pr(\chi/\Psi) = (Pr(\Phi_i/\Psi) * Pr(\chi/\Phi_i \wedge \Psi)).$$

Proof. Assume the hypotheses of the theorem. Then by the definition of joint exhaustiveness, $\Psi \models (\Phi_1 \vee \ldots \vee \Phi_n)$. By deductive logic, $(\Phi_1 \vee \ldots \vee \Phi_n) \models (\chi \equiv ((\Phi_1 \wedge \chi) \vee \ldots \vee (\Phi_n \wedge \chi))$, whence again by deductive logic, $\Psi \models \chi \equiv ((\Phi_1 \wedge \chi) \vee \ldots \vee (\Phi_n \wedge \chi))$. By our Lemma, $Pr(\chi/\Psi) = Pr((\Phi_1 \wedge \chi) \vee \ldots \vee (\Phi_n \wedge \chi)/\Psi)$. From our original assumption and deductive logic, $\Phi_1 \wedge \chi, \ldots, \Phi_n \wedge \chi$ are mutually exclusive on Ψ. From the original assumption, Axiom 1 gives that $Pr(\chi/\Psi) = (\Phi_i \wedge \chi/\Psi)$ Axiom 2 gives that $Pr(\chi/\Psi) = Pr(\Phi_i/\Psi)\ x\ \chi/(\Phi_i \wedge \Psi)$. ■

This theorem begins to show how the foregoing results are supposed to benefit a theory of induction, rather than simply show us the reasoning behind the theory of statistics. The key element here is the 'stacking' of probabilities. Given that if a certain condition or context of possibility Ψ obtains, one of $\Phi_1, \ldots, \Phi_n$ obtains by definition of mutual exclusivity and joint exhaustion. Further, there is a given probability that χ will obtain, and this probability might be different depending upon which Φ_i actually

obtains. For each Φ_i which might obtain, the probability of that Φ_i obtaining must be multiplied by the probability that χ might obtain *given* that Φ_i obtains (and also given Ψ, but this is beside the point). We now have the probability of χ *relative* to the given Φ_i. However, as some other $\Phi_i$might obtain, and its probability of χ given it may be different, we must calculate the sum of all the probabilities of χ *relative* to each given Φ.

So $Pr(\chi/\Psi)$ really equals

$$(Pr(\Phi_1/\Psi) * Pr(\chi/\Phi_1 \wedge \Psi) \vee \ldots \vee Pr(\Phi_n/\Psi) * Pr(\chi/\Phi_n \wedge \Psi))$$

Suppose we wish to determine the probability of a certain event in the world. Let us call this χ. Further suppose that we have a certain number of theories about the world, $\Phi_1, \ldots, \Phi_n$. We do not know which of these theories actually obtains, but let us also stretch our imaginations and suppose that they are mutually exclusive and together exhaust all the possible ways that a world could be.

Event χ is highly probable according to some of these theories and highly improbable according to others. However, these theories themselves are also more or less probable. For example, some people have theories about the world in which seeing Elvis is a probable occurrence. These theories, however, are themselves highly improbable. So a high probability *given* a theory may be radically discounted by the improbability *of* that theory. The exact mechanics of this discounting (or, conversely, of an accentuating) is given via this theorem. So, to determine the probability of seeing Elvis we first find the probability of the individual theory's obtaining and then multiply by the probability that our target occurrence (in this case, seeing Elvis) will obtain granted that theory. What we will have then is a number of probabilities of seeing Elvis, each being relative to a given theory. Add these all up, and you have the probability of seeing Elvis.

THEOREM 14. *If $\Phi_1, \ldots, \Phi_n$ are mutually exclusive and jointly exhaustive on condition χ, then for* j = *1, …,* n*:*

$$Pr(\Phi_j/\chi_i \wedge \Psi) = \frac{Pr(\Phi_j/\Psi) * Pr(\chi/\Phi_j \wedge \Psi)}{\sum_{j=1}^{n} Pr(\Phi_j/\Psi) * Pr(\chi_h/\Phi_j \vee \Psi)}$$

Proof. Assume the hypotheses of the theorem. It is also assumed that $Pr(\chi/\Psi) \geq 0$. By deductive logic $(\chi \wedge \Phi_j) \models (\Phi_j \wedge \chi)$ of each $j = 1, \ldots, n$. Then, by Theorem 5, $Pr(\chi \wedge \Phi_j/\Psi) = Pr(\Phi_i \wedge \chi/\Psi)$. Axiom 2 gives $Pr(\chi/\Psi) *\ Pr(\Phi_j/\chi \wedge \Psi) = Pr(\Phi_j/\Psi) *\ Pr(\chi/\Phi_j \wedge \Psi)$. And arithmetic gives

$$Pr(\Phi_i/\chi \wedge \Psi) = \frac{Pr(\Phi_j/\Psi) * Pr(\chi/\Phi_j \wedge \Psi)}{Pr(\chi/\Psi)}$$

Given our original assumption, Theorem 13 implies

$$Pr(\Phi_j/\chi_n \wedge \Psi) = \frac{Pr(\Phi_j/\Psi) * Pr(\chi/\Phi_j \wedge \Psi)}{\sum_{j=1}^{n} Pr(\Phi_j/\Psi) * Pr(\chi/\Phi_j \wedge \Psi)}$$

■

Let there be three closed bins, indistinguishable in appearance. In one there are eight green and two red apples; in another there are five of each kind; and in the third there are four green and six red apples. A bin is picked at random and three apples (with replacement) are drawn. The selection is subject to three hypotheses.

H1. Eight green and two red
H2. Five of each
H3. Four green and six red.

Suppose now that we have the further fact E that in three draws (with replacement) all the apples were green.

We can now determine the probabilities of *H1*, *H2* and *H3* in the light of evidence E. Theorem 8 requires that these probabilities sum to 1; so the probability of one of the hypotheses on E is got by subtracting the sum of the probabilities of the other hypotheses from 1:

$$\begin{aligned} &Pr(H1/D) = \\ &= \frac{Pr(H1) * (Pr(E/H1))}{Pr(H1)Pr(E/H1) + Pr(H2) * Pr(E/H2) + Pr(H3) * Pr(E/H3)} \\ &= \frac{(1/3) * (8/10)^3}{(1/3)(8/10)^3 + (1/3) * (5/10)^3 + (1/3) * (4/10)^3} \\ &= \frac{512}{701} \\ &= 0.7304 \end{aligned}$$

Similarly, $Pr(\mathrm{H2}/E)$

$$= \frac{(1/3) * (5/10)^3}{\frac{701}{3000}} = 0.1783$$

So, $Pr(H3/E) = 1 - (0.73 + 0.178) = 0.092$. As expected, the new evidence D raised the probability of $H1$ (from 0.33 to 0.73) and decreased the probability of $H3$ (from .33 to 0.092). Hence it appears that we can say that E confirms $H1$ and disconfirms $H3$.

Bayes' Theorem gives us occasion to return to the distinction between *prior* and *posterior* probabilities. The prior probability of an event is the value it has before the factoring in of new evidence. Its posterior probability is the value it receives after the new evidence is taken into account. More generally,

> $Pr(\Phi_i/\Psi)$ is the prior probability of Φ_i

and

> $Pr(\Phi_i/\chi \wedge \Psi)$ is the posterior probability of Φ_i (relative to evidence χ).

The values of χ for each of the hypotheses in question are the *inverse probabilities* of the hypotheses' posterior probabilities, viz., $Pr(\chi/\Phi_i \wedge \Psi)$. These inverse probabilities are often called "likelihoods" in a technical, not the generically intuitive sense of the word. As can be seen, Bayes' Theorem is a method of finding the posterior probabilities of a set of mutually exclusive and jointly exhaustive hypotheses from their prior probabilities and their likelihoods. Its relevance to actual reasoning is also now apparent, as is a limitation. The limitation is that often, though not always, Bayes' Theorem will not apply to situations in which precise probability values cannot believably be assigned.

The significant predecessor of Bayes' Theorem is obviously Theorem 13, and is the most important when the question of the role of confirmation arises. For this reason, it is the culmination of the **ProbCal** exposition to date which deals with the machinery of the probability calculus.

As in Theorem 13, with Bayes' Theorem the key concept is that of the 'stacking' of probabilities. We suppose that there is a background. Then there are a range of $\Phi_1, \ldots, \Phi_n$, each of which have a characteristic probability on Ψ. Obviously then, the probability of χ's obtaining is affected by which Φ_i actually obtains. The issue of this theorem, though, is given the occurrence of χ, what is the probability of a given Φ_i obtaining?

The theorem says that we may calculate the probability of a given Φ_i's obtaining on conditions χ and Ψ in the following manner:

1. (a) Find the probability of the target Φ_i simply given Ψ.
   (b) Find the probability of χ should both the particular target Φ_i *and* Ψ obtain.
   (c) Find the product of 1(a) and 1(b).
2. (a) For *each* Φ_i, find its probability simply given Ψ.
   (b) For *each* Φ_i, find the probability of χ should both that Φ_i and χ obtain.
   (c) Find the product of 2(a) and 2(b), for *each* Φ_i.
   (d) Add up all the 2(c)'s.
3. (a) Divide 1(c) by 2(d).

In essence what we're doing here is taking the probability of χ relative to the particular target Φ_i and dividing it by the probability of χ given all the Φ's. This should tell us which Φ most probably obtains.

Suppose, as an example, that we have three boxes with the following contents:

Box 1: 3 green marbles, 2 black marbles, 5 white marbles.

Box 2: 0 green marbles, 0 black marbles, 10 white marbles.

Box 3: 7 green marbles, 8 black marbles, 5 white marbles.

The states of affairs in boxes 1, 2 and 3 will be referred to as Φ_1, Φ_2, Φ_3 respectively. Now let χ be the occurrence of our drawing a white marble. To find the overall probability of drawing a white marble without knowing which box we have drawn from, we apply theorem 12.

> $Pr(\Phi_i/\Psi)$ of any of Φ_1, Φ_2, Φ_3 is $1/3$, as we may have reached into any one of the three boxes.
> $Pr(\chi/\Phi_1 \wedge \Psi)$ is $5/10$ or $1/2$, as $1/2$ of the marbles in the box are white.
> $Pr(\chi/\Phi_2 \wedge \Psi)$ is 1, as all the marble in box 2 are white.
> $Pr(\chi/\Phi_3 \wedge \Psi)$ is $5/20$ or $1/4$, as $1/4$ of the marbles in box 3 are white.

So

$$\begin{aligned}
&\textstyle\sum_{i=1}^{3} Pr(\Phi_i/\Psi) * Pr(\chi/\Phi_i \wedge \Psi) \\
&\quad = (Pr(\Phi_1/\Psi) * Pr(\chi/\Phi_1 \wedge \Psi)) + \\
&\qquad (Pr(\Phi_2/\Psi) * Pr(\chi/\Phi_2 \wedge \Psi)) + \\
&\qquad (Pr(\Phi_3/\Psi) * Pr(\chi/\Phi_3 \wedge \Psi)) \\
&\quad = (1/3 * 1/2) + (1/3 * 1) + (1/3 * 1/4) \\
&\quad = 1/6 + 1/3 + 1/12 \\
&\quad = 7/12
\end{aligned}$$

Thus $Pr(\chi/\Psi) = 7/12$, which is the denominator of our new formula. We may use this to find the probability of that marble having come from a given box.

(a) Let us now calculate the probability of the marble having come from box 1. Our formula tells us that:

$$Pr(\Phi_1/\chi \wedge \Psi) = \frac{Pr(\Phi_1/\Psi) * Pr(\chi/\Phi_1 \wedge \Psi)}{\sum_{i=1}^{3} Pr(\Phi_i/\Psi) * Pr(\chi/\Phi_i \wedge \Psi)}$$

(b) But we also have already calculated these numbers just previously:

$$Pr(\Phi_1/\chi \wedge \Psi) = \frac{\frac{1}{3} * \frac{1}{2}}{\frac{7}{12}} = \frac{\frac{1}{6}}{\frac{7}{12}} = \frac{2}{7}$$

Similarly.

$$Pr(\Phi_2/\chi \wedge \Psi) = \frac{\frac{1}{3} * 1}{\frac{7}{12}} = \frac{\frac{1}{3}}{\frac{7}{12}} = \frac{12}{21} = \frac{4}{7}$$

(c) And

$$Pr(\Phi_3/\chi \wedge \Psi) = \frac{\frac{1}{3} * \frac{1}{4}}{\frac{7}{12}} = \frac{\frac{1}{12}}{\frac{7}{12}} = \frac{1}{7}$$

There are just a few additional things to note. As a check, we may see that the probabilities sum to 1, which is simply the probability that the marble came from one of our boxes and did not just popout of thin air. A trivial point, this is a quick check on our calculations. Secondly, even though there were just as many marbles in box 3 as in box 1, it is twice as probable that we had reached into box 1, as it was twice as likely that pulling a marble from box 1 would result in the drawing of a white marble (once again, a trivial point).

A non-trivial point is that we have come up with a theory which seems to be couched in deductive terms; a theory which seems to give us a means of adjudicating between rival hypotheses. If we pluck a few more marbles from the same box and they keep coming up white, we may apply this theorem and axiom 2 until the probability of our pulling marbles from either box 1 or 3 becomes incredibly low. Even if 5 white marbles were successively plucked from the same box, the stout skeptic could maintain that it is *possible* that we're pulling from box 1 or 3; but it seems clear that such a person might want to steer clear of Las Vegas. A pertinent question now might be, "Could this sort of theory ground our theories of confirmation?"

5 DUTCH BOOKS

Suppose that a degree of belief in a statement is measured by the betting odds one would be ready to accept on that statement. The Dutch Book theorem[10], to which we now turn, asserts that a failure to set one's degrees of belief, or subjective probabilities, in fullfuilment of the theorems of **ProbCal** exposes one to a betting strategy that guarantees that one loses money no matter whether one's belief is true or false. Coherence may be seen, therefore, as belief-formation that evades the Dutch Book; i.e., that satisfies **ProbCal**. Thus, Dutch Book theorem is a motivator of *coherence*.

[10]See here Gustason [1994].

Some writers (e.g., van Fraassen) also see it as a motivator of the Reflection Principle. Informally, the Reflection Principle provides that if an agent has reason to believe that his degree of belief in statement Φ will have the value r in the future, then he is rationally required to set the value of Φ at r *now*, never mind the particularities of psychological doubt or certainty that now pertain. Not everyone will see the Reflection Principle as epistemologically obvious. Perhaps there is reason to wonder about the philosophical cogency of Dutch Books arguments. (See here Earman [1992, chapter 2].) Let us see.

It is perfectly natural to wonder whether the axioms of **ProbCal** are actually true and, relatedly, whether we are in fact justified in conforming our conception of probability to them. One way of approaching these questions is to consider whether conformity to the axioms would constitute a *bet* as *rational*.

A bet is a bet on a proposition Φ at odds x:y against the truth of Φ. Bearing in mind that probability is conditional probability, it is better to conceive of bets as bets on pairs of propositions $\langle \Phi, \Psi \rangle$ (i.e., $\langle \Phi$, given $\Psi \rangle$); where it is understood that Ψ conveys such evidence as there may be for Φ. Using the dollar as our exemplary currency, a *bet* on $\langle \Phi, \Psi \rangle$ at odds x:y is one in which the wagerer *wins* \$x if $\langle \ulcorner \Phi \wedge \Psi \urcorner \rangle$ is true and *loses* \$y if $\langle \ulcorner \neg \Phi \wedge \Psi \urcorner \rangle$ is true. If $\ulcorner \neg \Psi \urcorner$ is the case, the bet is "off". The *stakes* of such a bet are x+y, and the *betting quotient*, *BQ*, of $\langle \Phi, \Psi \rangle$ is y/x+y). Thus *BQ* $\langle \Phi, \Psi \rangle$ is a *ratio of loss to stakes.*

A wager system is an arrangement between two agents $\mathbf{A}_1$ and $\mathbf{A}_2$ who have come to an agreement on: the odds for each bet, the stakes for each, and who bets on or against $\langle \Phi, \Psi \rangle$.

A bettor, A_1, *makes a book* against a bettor A_2, where A_2 is a set of "clients", when on a number of $\langle \Phi, \Psi \rangle$, he sets the odds so as to attract the custom of A_2 and yet guarantees himself a profit no matter what the outcome is. Since typically A_2 is a large clientele, certainly larger than a single individual, it is possible for A_1 to make a book against his collective clientele without making a book against any single client. However, this last thing *is* possible, and when A_1 brings it off against an individual, A is said to have made a *Dutch Book* against that individual.

If a bettor A_1 sets out to make a Dutch Book against some individual opponent A_2, his success depends on how he, together with A_2, manages to contrive an appropriate wager-system. For an examination of such things we will need the concept of *expected value*, with the help of which we will be able to define a *fair* bet.

Expected value is a central notion in *decision theory.* It is obvious that often, indeed typically, a human agent acts in a situation in a way that represents only one of a number of possible action-alternatives in that situation. Also typical is that the agent chooses the action-alternative in question without knowing for certain the consequences either of the action he chose or of

the others he didn't choose. Action in such circumstances is action *under uncertainty.* Agents acting under uncertainty will seek mitigation (though seldom removal) of the uncertainty in two ways. They will attempt to estimate the probable consequences of the action-alternatives; and they will also attempt to determine the extent to which those consequences, if they occurred, would answer to the agents' interests. In other words, agents will consider the *values* of the probable outcomes of the various act-alternatives in question.

Let a and c be variables ranging over *acts* and *consequences* respectively. Let the schematic letters A and C represent specific acts and consequences respectively. We also stipulate that a, c, A and C may occur with subscripts. It is possible, then, to define the expected value of a on Ψ by the Principle of Expected Value.

> **PrExVal**: If $\Phi_1, \ldots, \Phi_n$ are mutually exclusive and jointly exhaustive on condition Ψ, and c_i is a consequence of act a that obtains if Φ_i holds true, then $\mathbf{ExVal}(a/\Psi) = \sum_{i=1}^{n} (Pr(\Phi_i/a \wedge \Psi) * \mathbf{Val}(c_i))$.

Accordingly, β is a *fair bet* on $\langle \Phi, \Psi \rangle$ at odds $x{:}y$ if and only if $\mathbf{ExVal}\ (\beta/\Psi) = Pr(\Phi/\Psi)x + Pr(\neg\Phi/\Psi)$-y $= 0$

(It is assumed that the action of betting β does not influence the probability of Φ or Ψ.) With the definition at hand, and with the help of **Thm2** and some arithmetic, it is easy to show that a fair bet on $\langle \Phi, \Psi \rangle$ is one in which the *probability* of $\langle \Phi, \Psi \rangle$ and the *betting quotient* of $\langle \Phi, \Psi \rangle$ are equal. Similarly, the betting quotient is fair just in case it equals the probability.

We may now say that a wagering system W has been *established from* a set of betting quotients when, for each bet in W, the stakes have been agreed by A_1 and A_2, as well as which bettor will make which bet. A Dutch Book is a W in which all outcomes favour one of A_1, A_2, and are against the other. The one agent has made a Dutch Book *against* the other.

We now record an important negative condition on Dutch Books.

> K1: if σ is a set of fair betting quotients (recall, BQ$\langle \Phi, \Psi \rangle$ is fair iff $BQ\langle \Phi, \Psi \rangle = Pr(\Phi/\Psi)$ for all $\langle \Phi, \Psi \rangle$), then *no* Dutch Book can be established from σ if and only if it satisfies the axioms of **ProbCal**.

$K1$ is a biconditional, hence the conjunction of two conditional statements K_1a and K_1b. We turn first to K_1a, the claim that if those probabilities *don't* conform to the axioms of **ProbCal**, then a Dutch Book can be established from σ.

To establish this result, we will need to consider our five axioms. We shall consider each axiom in order, in light of an example that violates it. We shall try to show that in each such case a Dutch Book is establishable

from the relevant σ. For ease of exposition we will consider variations of the axioms in which conjuncts and disjuncts are restricted to just two. Our results easily generalize to longer compounds.

AXIOM 1 If Φ and Ψ are mutually exclusive on χ then $Pr(\Phi \vee \Psi/\chi) = Pr(\Phi/\chi) + Pr(\Psi/\chi)$.

We can see that if Φ and Ψ are mutually exclusive on χ, then what we may call the *sum* of bets on $\langle\Phi,\chi\rangle$ and $\langle\Psi,\chi\rangle$ at equal stakes is itself a bet on $\langle\Phi \vee \Psi,\chi\rangle$ at those same stakes. To show this, consider any Φ and Ψ mutually exclusive on χ and let β_1 be a bet on $\langle\Phi,\chi\rangle$ at odds $x : y$, and β_2 a bet on $\langle\Psi,\chi\rangle$ at odds $z : w$. We put it that x+y=z+w. Given the mutual exclusiveness of Φ and Ψ, there are just three cases to consider:

		β_1	β_2	**Sum** (β_1,β_2)
Case one:	$\langle\Phi,\chi\rangle\langle\neg\Psi,\chi\rangle$	x	-w	x-w
Case two:	$\langle\neg\Phi,\chi\rangle\langle\Psi,\chi\rangle$	-y	z	z-y
Case three:	$\langle\neg\Phi,\chi\rangle\langle\neg\Psi,\chi\rangle$	-y	-w	-(y+w)

Consider the column headed by "Sum(β_1, β_2)". If the values in the two top rows are identical, then the *sum* is in its own right a winning bet if either Φ or Ψ obtains and a loser if neither does. In this situation the column describes a bet, β_3 on $\langle\Phi \vee \Psi,\chi\rangle$ which wins x-w dollars (i.e., $z - y$ dollars) in cases Case one and Case two and loses $y + w$ dollars in case Case three. Now $x + y = z + w$, so $x - w = z - y$. Because β_1 is a bet on $\langle\Phi,\chi\rangle$ at x:y and β_2 is a bet on $\langle\Psi,\chi\rangle$ at $z : w$, then sum(β_1, β_2) is a bet β_3 on $\langle\Phi \vee \Psi,\chi\rangle$ at $x-w : y+w$. This means that β_3's stakes are $(x-w)+(y+w)$ which equals $x + y$, which equals z+w.

Imagine that A_1 and A_2 are betting on the next Bluejays game. Let V be the proposition that the Bluejays win, and T the proposition that they tie. Let E be whatever evidence is available at the time of the bet. Clearly V and T are mutually exclusive on E. A_1 reckons that $Pr(V/E) = \frac{3}{8}$ and $Pr(T/E) = 1/4$. From A_1's point of view these are fair bets:

β_1: Bet on $\langle V, E\rangle$ at 5:3
β_2: Bet on $\langle T, E\rangle$ at 3:1

From what we have proved in the preceding paragraph, we have it that sum (β_1, β_2) is a bet on $\langle V \vee T, E\rangle$ at the same stakes. Axiom 1 tells us that $Pr(V \vee T/E) = 5/8$. So on odds 3:5 this is a fair bet.

We might easily imagine that at this point A_1 trips up. Instead of reckoning $Pr(V \vee T/E)$ to be 5/8, let us suppose that he puts it at 1/2. Accordingly A_1 considers the bet fair at odds of 1:1. A_1 would also consider fair a bet against a win or a tie. This is a bet on $\langle\neg(V \vee T), E\rangle$ at odds of 1:1. Call this bet β_3.

Having violated Axiom 1 in the manner indicated, we must try on behalf of A_1's opponent A_2, to make a Dutch Book against A_1. This is easily done.

A_2 need do no more than accept each of β_1, β_2 and β_3 *at the same stakes.* Suppose the stakes are \$8.00, then all three together establishes a Dutch Book against A_1:

		β_1	β_2	β_3	**Sum**β_1,β_2,β_3
Case one:	$\langle V, E\rangle\langle\neg T, E\rangle\langle V \vee T, E\rangle >$	5	-2	-4	-1
Case two:	$\langle\neg V, E\rangle\langle T, E\rangle\langle V \vee T, E\rangle$	-3	6	-4	-1
Case three:	$\langle\neg V, E\rangle\langle\neg T, E\rangle\langle\neg(V \vee T), E\rangle$	-3	-2	4	-1

No matter the bet, A_1 is always the net loser by a dollar. Had A_1 attended to **ProbCal** all would have been well. He would have put $\langle V \vee T, E\rangle$ at 5/8 and $\langle\neg(V \vee T), E\rangle$ at 3/8, giving payoffs for β_3 of -\$3.00 for cases one and two and \$5.00 for case three. The sum (β_1, β_2, β_3) would have been 0 for all cases, thus precluding a Dutch Book.

Even so, A_1 might have violated Axiom 1 in a different way, by assigning to $\langle V \vee T, E\rangle$ a value in excess of 5/8, say 3/4. Suppose that $\mathbf{A}_1$ could be got to bet against $\langle V, E\rangle$ and $\langle T, E\rangle$ but on the disjunction, i.e., $\langle V \vee T, E\rangle$. Then there are three bets that A_1 would now consider fair:

β_4: Bet on $\langle\neg V, E\rangle$ at 3:5
β_5: Bet on $\langle\neg T, E\rangle$ at 1:3
β_6: Bet on $\langle V \vee T, E\rangle$ at 1:3

A_2 has a Dutch Book against A_1:

	β_4	β_5	β_6	**Sum**β_4, β_5, β_6
Case one:	-5	2	2	-1
Case two:	3	-6	2	-1
Case three:	3	2	-6	-1

Our examples generalize:

> If Φ and Ψ are mutually exclusive on χ and a bettor (i) gives that $Pr(\Phi/\chi) = x$, $Pr(\Psi/\chi) = y$ and $Pr(\Phi \vee \Psi/\chi) = z \neq x + y$, and (ii) accepts fair bets on each set of objects at equal stakes, then a Dutch Book is makeable against this bettor if he also accepts bets *on* $\langle\Phi, \chi\rangle$, *on* $\langle\Psi, \chi\rangle$ and *against* $\langle\Phi \vee \Psi, \chi\rangle$ if $z < x+y$, and *against* $\langle\Phi, \chi\rangle$, *against* $\langle\Psi, \chi\rangle$ and *on* $\langle\Phi \vee \Psi, \chi\rangle$, if $z > x + y$.

Thus, noncompliance with Axiom 1 exposes one to a Dutch Book.

We now turn to violations of Axiom 2.

AXIOM 2 $Pr(\Phi \wedge \Psi/\chi) = Pr(\Phi/\chi) * Pr(\Psi/\Phi \wedge \chi)$.

Let F be the statement that the Bluejays star pitcher has an injury-free season, and P be the statement that the Bluejays win their division. Let E be all the evidence available to agents who might want to wager on these possibilities. We consider bets on: $\langle F, E\rangle$, $\langle P, F \wedge E\rangle$ and $\langle F \wedge P, E\rangle$. Since

the second bet is conditional on the star pitcher's avoiding injury, if he is injured the bet is off.

Suppose that A_1 reckons that $Pr(F, E) = 3/5$ and $Pr(P/F \wedge E) = 5/8$. By Axiom 2, $Pr(F \wedge P/E) = 3/8$, but we will imagine that A_1 ignores it, or is unaware of it, and assigns the value 1/4. Then A_2 can get A_1 to accept the following bets as fair

β_1: Bet on $\langle F, E\rangle$ at 2:3
β_2: Bet on $\langle P, F \wedge E\rangle$ at 3:5
β_3: Bet on $\langle \neg(F \wedge P), E\rangle$ at 1:3

and if A_1 also accepts the following stakes: x for both β_2 and β_3, and xy for β_1, where $y = Pr(P/F \wedge E)$, then a Dutch Book can be made against A_1. If A_1 accepts \$8.00 stakes for β_2 and β_3, the stakes for β_1 are $(5/8)8 =$ \$5.00. For these stakes, there are three outcomes in each of which A_1 has a net loss:

		β_1	β_2	β_3	**Sum**β_1,β_2,β_3
Case one:	$\langle \neg F, E\rangle$,—,$\langle \neg(F \wedge P), E\rangle$	-3	—	2	-1
Case two:	$\langle F, E\rangle\langle \neg P, F \wedge E\rangle\langle \neg(F \wedge P), E\rangle$	2	-5	2	-1
Case three:	$\langle F, E\rangle\langle P, F \wedge E\rangle\langle F \wedge P, E\rangle$	2	3	-6	-1

Of course, here too $\mathbf{A}_1$ might have reckoned the value of $Pr(F \wedge P/E)$ to be greater than 3/8. With the stakes the same as before, if A_2 can induce A_1 to bet *against* $\langle F, E\rangle$, *against* $\langle P, F \wedge E\rangle$ and *on* $\langle F \wedge P, E\rangle$, a Dutch Book is makeable against him.

We again generalize our result.

> If a wagerer reckons that $Pr(\Phi/\chi) = x$, $Pr(\Psi/\Phi \wedge \chi) = y$ and $Pr(\Phi \wedge \Psi/\chi) = z{\neq}xy$, and (ii) accepts fair bets in relation to $\langle \Phi, \Psi\rangle$ at stake w for $\langle \Psi, \Phi \wedge \chi\rangle$ and for $\langle \Phi \wedge \Psi, \chi\rangle$ and at stakes yw for $\langle \Phi, \chi\rangle >$ then a Dutch Book can be made against this wagerer if he accepts bets *on* $\langle \Phi, \chi\rangle$, *on* $\langle \Psi, \Phi \wedge \chi\rangle$ and *against* $\langle \Phi \wedge \Psi, \chi\rangle$ if $z < xy$, and *against* $\langle \Phi, \chi\rangle$, *against* $\langle \Psi, \Phi \wedge \chi\rangle$ and *on* $\langle \Phi \wedge \Psi/\chi\rangle$ if $> xy$.

We take it as obvious that $K1$(a) holds for the remaining, entirely obvious, axioms of **ProbCal**. So, $K1$(a) is verified.

$K1$(b) is the other half of the biconditional $K1$. It asserts that if the probability values from which σ was derived obey the axioms of **ProbCal**, then no Dutch Book can be established from σ. We begin with the assumption that if σ is a series of fair bets, their sum will also be fair. Suppose that Φ and Ψ are independent of χ, and let us set $Pr(\Phi/\chi)$ at 1/2 and $Pr(\Psi/\chi)$ at 1/4. Then a bet β_1 on $\langle \Phi, \chi\rangle$ is fair at odds 1:1, and a bet β_2 on $\langle \Psi, \chi\rangle$ is fair at odds 3:1. A wagering system W arises from the set of betting quotients if a bettor A_1 undertakes to bet β_1 and β_2 and A_1 and A_2 agree stakes of \$2.00 for β_1 and \$4.00 for β_2. The betting setup is this:

		β_1	β_2	**Sum** (β_1, β_2)
Case one:	$\langle\Phi,\chi\rangle\langle\Psi\chi\rangle$	1	3	4
Case two:	$\langle\Phi,\chi\rangle\langle\neg\Psi,\chi\rangle$	1	-1	0
Case three:	$jset\neg\Phi,\chi\langle\Psi,\chi\rangle$	-1	3	2
Case four:	$\langle\neg\Phi,\chi\rangle\langle\neg\Psi,\chi\rangle$	-1	-1	-2

We ask, what is the expected value of the sum of these bets? β_1 and β_2 are both fair; so their expected values is 0. The same is true of their sum. That is,

ExVal: (sum$(\beta_1, \beta_2)/\chi$)
$=Pr(\Phi\wedge\Psi/\chi)*(\$4)+Pr(\Phi\wedge\neg\Psi/\chi)*(0)+Pr(\neg\Phi\wedge\Psi/\chi)*(\$2)+Pr(\neg\Phi\wedge \neg\Psi/\chi)*(\text{-}\$2)$
$= (1/8)*(4) + (1/8)*(2) + (3/8)*(\text{-}2)$
$= 1/2 + 1/4 - 3/4 = 0$

It is easy to see that any wagering system established from the betting quotients in question will have an **ExVal** of 0, never mind how the stakes are altered and never mind who makes that bet. Thus if each bet of a class of bets is fair, so too is their sum (and the sum of such sums, too). But the principle fails in reverse. A sum of bets might be fair even though the individual bets are not. Consider, for example, two bets whose **ExpVal**s are $4.00 and -$4.00. They are not fair bets even though their sum is fair.

As we have seen, determinations of **ExpVal** presuppose the axioms and theorems of **ProbCal**. This makes it attractive to suppose that if the probability values underwriting a set of fair betting quotients obey the **ProbCal** axioms, then the sum of fair bets with respect to those quotients will also be fair. A wagering system is, so to say, nothing but a *sum of bets*. Hence any wagering system establishable from the set of fair betting quotients will have an **ExpVal** of 0. The proof of this is furnished by Shimony [1955] and will be omitted here in the interest of space.

Given Shimony's proof, we are entitled to say that if the inductive probabilities that fix a set of fair betting quotients satisfy the **ProbCal** axioms, then each wagering system establishable there-from is fair. If we concede that a Dutch Book is an unfair wagering system, then we may take it that K1(b) is verified.

The theorist whose interest lies chiefly in practical reasoning (in one of its meanings), that is, in those inferences and arguments that arise in the course of our everyday thinking about everyday things, may find himself less than enchanted by the present defence of the **ProbCal** axioms. If his disenchantment arises from the conviction that wagering systems are an unrepresentative sample of practical reasoning the probability theorist has an attractive reply to such criticism. On the other hand, if the practical reasoner has in mind a further technical result in the theory of wagering systems itself, his disenchantment is more difficult to brush aside. We take up these points in order.

As for the first, it is natural to think that wagering strategies are untypical of inferences as such. Against this, it may be said that most actions are gambles of a sort. To the extent that they are at all rational they involve the interplay of probabilities and expected values. This appears to be so even for quite ordinary actions and for such reflection as may attend upon them, whether it is crossing Fifth Avenue at high noon or driving your car to Banff (or anywhere else). What seems to pull the fangs of the first criticism of the defence of the **ProbCal** axioms is this: A Dutch Book will obtain *whenever* an agent, Daria say, invests a proposition or a pair of propositions with a *degree of belief* which is inconsistent with the degree of belief assigned to certain other propositions. More generally,

> **ConBel**: Compliance with the **ProbCal** axioms is a condition on the self-consistency of an agent's degrees of belief.

To the extent that the inference theorist, the theorist of argument, or the theorist of practical reasoning has a standing in determining what influence, if any, an agent's internal consistency has on his arguments, his inferences or anything else he reasons about, it is helpful to know that compliance with these axioms spares a reasoner the nuisance of falling into inconsistency by way of Dutch Book susceptibilities.

Even so, all is not well, as an examination of the critic's second objections makes clear. Although consistency with the **ProbCal** axioms arguably precludes the emergence of a Dutch Book, it does not rule out a *partial book*. A partial book is a wagering system in which one of the wagerers never wins but does not always lose. Consider Axiom 3: If $\Phi \models \Psi$ then $Pr(\Psi/\Phi) = 1$. Consider now an A and a B such that $Pr(\mathrm{A/B}) = 1$. Then a fair bet (so to speak) on $\langle A, B\rangle$ will be at odds 0:1. If $B \models A$ then you cannot possibly lose this bet, but there is no profit in winning it. Let us slightly change the present example. It is logically possible that A is false and B true even though $Pr(\mathrm{A/B}) = 1$. Then a fair bet on $\langle A, B\rangle$ gives a partial book:

		Bet on $\langle A, B\rangle$
Case one:	$\langle A, B\rangle$	0
Case two:	$\langle \neg A, B\rangle$	-1

The axioms of **ProbCal** are powerless to block such books. But if we were to proclaim the *converse* of Axiom 3, viz.,

> If $Pr(\Psi/\Phi) = 1$, then $\Phi \models \Psi$

we would shut down the very assumption that permits the partial book to arise. Accordingly, the preclusion of both Dutch and practical books require fidelity to the **ProbCal** axioms supplemented by the converse of Axiom 3. The converse of Axiom 3 is fully independent of the **ProbCal** axioms, i.e.,

neither it, nor its negation, is derivable from them. Moreover, it doesn't seem to be true, and that is a problem.

6 BAYES' THEOREM AND BAYESIAN INFERENCE

One of the uses to which we put our inferences is the revision or updating of our belief-sets. Sometimes it suffices to use purely deductive procedures, but typically not. One of the tasks of inductive logic is to specify at least some of the non-demonstrative protocols for belief-revision.

A subsidiary, but no less important question, is whether **ProbCal** gives us any direct information about how such procedures work.

Bayesianism: "Bayesianism" is the name given to a rather widely held position by inductive logicians. In its most basic sense, as we have seen, it is a view which identifies inductive likelihood with the probability notion of the calculus of probability. But in another dominant use, it is a position according to which satisfaction of the belief-revisionist's first task involves giving an affirmative answer to our present question. A word of caution is necessary. A theory of belief-revision is not made Bayesian simply by its subscription to Bayes' theorem. Not every theory in which Bayes' theorem holds is a Bayesian theory. There is in fact considerable (and contentious) latitude given to the contemporary meaning of "Bayesianism" (Earman [1992]). Even so, there is considerable agreement about the following core idea. As a first approximation, a theory is Bayesian to the extent that

(a) it interprets probability subjectively

and

(b) sanctions the claim that belief-revision (or inference) is conditionalization.

We now turn to a fuller consideration of these claims.

The section on Dutch books is designed to show what happens when we undertake to bet on propositions which behave according to the axioms of **ProbCal**. The point of the doctrine of Dutch Books is to describe what attitude one should taken when confronted with a system which does in fact work according to the axioms of **ProbCal**.

The term "Dutch Book" comes from two sources: (1) During the early enlightenment there was intense economic rivalry between the Dutch and the English (actually between the Dutch and everyone else), and the English attempted to associate the Dutch with anything unsavory or unfair (hence Dutch uncle, Dutch courage, Dutch treat, etc.). (2) A "book" is an arrangement whereby a bet is arranged at odds favourable to one of the bettors (usually known as the bookmaker). A Dutch book is then an extreme

case of the more usual book, in that things have been arranged so that the bookmaker cannot lose.

The betting simile seems appropriate in as much as it parallels certain epistemic practices. Each and every one of us has what might be called a *belief system*. For example, Sarah believes that the law of gravity governs certain aspects of the relationship between her pen and the Earth. She believes that my coffee cup exists even when it is in the cupboard and she cannot see it, and so on. All of the beliefs that she holds are said to be held together in her *belief set*. It is also widely held that our beliefs admit of degrees; that is, that we may believe some things more than others. For example, Harry currently believes that his car has not been stolen. He would reply, if asked, as to the whereabouts of the car, that it is parked in his garage. Should he receive a phone call right now from a member of the RCMP advising him that the car was just now found to be in the hands of some joy-riders across town, he would not complain of police-mendacity. On the other hand, had someone told him that he had a magic 200 kg anvil which routinely floated unsupported in his living room, Harry would robustly disbelieve this statement. This is because Harry's degree of belief in his car's whereabouts is lower than his degree of belief about the ubiquitousness of gravity. In fact, it is not at all atypical to both believe a thing and its negation, in varying degrees.

The question brought up by our current investigation is how we ought to take this. If we believe that Φ, and evidence later shows that $\ulcorner \neg\Phi \urcorner$, we must certainly do some overhauling of our belief sets. If, however, we merely believe something like *probably* Φ, our belief sets may be fine, even in the face of overwhelming evidence of $\ulcorner \neg\Phi \urcorner$. For example, it seems all right if we believe to a high degree (i.e., that it is highly probable) that "the coin will not come up heads 10 times in a row", as the probability of this is rather low. Should 10 successive flips of a coin prove us "wrong", we do not have to revise our belief, since the belief was really that the occurrence mentioned would probably not happen; all the coin tosses showed was that the highly improbable still sometimes obtains. In other words, while the coin's yielding heads ten times successively *is* a disconfirmation of "the coin will not come up heads 10 times in a row", it is *not* a disconfirming instance of "it is very likely that the coin will not come up heads 10 times in a row". Our assertion is that the latter sentence has a degree of belief imbedded into it in the first four words, and this is what gets us out of trouble. So it seems that we might say that propositions admit of degrees of belief much as they admit of degrees of probability. The question is how much alike are these two concepts? Let us suppose for the moment that they work just alike. So we can see how a belief system which did work according to **ProbCal** would be constructed.

How much belief should we grant to a given proposition? How much belief is warranted? The answer is that it depends upon how much belief we invest

in other, related propositions. How it depends on these is dictated by the axioms of **ProbCal**. It would seem that, should a proposition Φ be almost certain, having perhaps a probability of .95, then we should be willing to lend that proposition *all* our belief. This is not as yet irrational (although we would be forced into a revision of our beliefs if that unlikely event with a probability of .05 actually came to pass), as we may well invest no belief whatever in $\ulcorner\neg\Phi\urcorner$. But if we wish to set our belief in $\ulcorner\neg\Phi\urcorner$ at .05 to "cover ourselves", while remaining wholly committed to Φ (that is believing in Φ to degree 1.0), this is tantamount to betting on Φ at odds 0:1 while betting on $\ulcorner\neg\Phi\urcorner$ at 19:1.

Dutch-style books arise whenever the odds given do not reflect the probabilities in question. If $\{\Phi, \Psi\}$ is a jointly exhaustive and mutually exclusive set of outcomes of a given situation, and if $Pr(\Phi)$ is .25 and $Pr(\Psi)$ is .75, and if Harry gives Sarah odds of 1:1 on Φ, then that means that Harry has set it up so that the expected value of Sarah's betting situation, supposing the bet to be an even \$100 is $(.25 * 100)+(.75 * \text{-}100) = 25 \text{ - } 75 = \text{-}50$. Of course, the value for Harry for this is \$50.00. Obviously, the bet has been skewed, and has been created this way by juggling the odds so that they do not fall into accord with the probabilities of the system. So then, when faced with a system which works probabilistically, one should be sure that the odds do not favour the other side.

7 ACTUAL VALUES

We now introduce a concept of Actual Value. This can be constructed to provide a distinction between the actual value of a situation, and the value that we expect that situation to have. The actual value of the situation, as is evidenced by the principle of expected value, is the probability of the event in question multiplied by the value of that occurrence. There is also a perceived value which is constructed as the product of the value of the occurrence and the perceived probability. What we need to be on the lookout for is the construction of a perceived value which is not the same as the actual value k. Divergence in the perceived probability of our given occurrence will lead us to put either too much stock in the given occurrence's happening, or too little.

As we see, there are two distinct ways in which we can go wrong in our formulation of degrees of belief in propositions. We might try to underestimate the chances of an occurrence, so as to create more value for the situation than is fairly attributed to it, or we might try to undervalue the situation, in which case an interesting consequence arises. Suppose instead of getting the thing dead right, we are somewhat skeptical about our assessment of the probability of the occurrence, and thus conservative in the arrangement of our bets. For example, consider a bet on Φ where $Pr(\Phi)$

= .75. The fair bet would be at odds 1:3 to reflect the 3/4 chance of winning. One might instead set the odds at 2:3 so as to increase the expected value of the situation to .75. This would correspond, in our belief model, to believing in Φ to degree 0.6. There might be some difficulty in attracting human bettors to these odds, but it seems that with belief, we may set our degrees of belief in any way we see fit, and that reality is simply forced to take us up on the wager. This is where the analogy breaks down.

The divergence from the betting analogy and the explanation of belief systems requires a realization that when we're arranging our belief systems we make a bet on both propositions. A subtle problem arises when we constantly underestimate the odds of the propositions in which we invest our interest. This seems the backbone of conservatism of all sorts, and as such we may have to fall back on our intuitions to judge this. That is, we invest both Φ and $\ulcorner\neg\Phi\urcorner$ with a degree of belief. If Harry believes Φ to degree 0.6, then he also believes in $\ulcorner\neg\Phi\urcorner$ to degree 0.4. So, while his bets on Φ seem conservative, his bets on $\ulcorner\neg\Phi\urcorner$ are over-optimistic.

In order that we may have a better pool of information from which to base our evaluation of the mechanics of **ProbCal** as an inductive logic, it will prove useful to look at the concept of conditionalization. Conditionalization is the process of finding what a probability would be once we acknowledge the realization of some given proposition. Let us take a look at what the formal machinery actually does, and then we might be able to interpret it satisfactorily.

Suppose we have some new evidence statement e. Before this statement presented itself, we had some measure of the probability of e, if only through simple enumeration. This must now change, as we now suppose that the situation reported by e actually does obtain. How does this affect the probabilities of the rest of our propositions, for example, the proposition that p? There will have been a prior probability of its conjunction with **e** (either through simple enumeration, or via whatever paths we may desire to posit). That is, there must have been an old $Pr(p \wedge e)$. But we now know that e actually obtains. So we want to "factor out" the old $Pr(e)$ from the old $Pr(p \wedge e)$, which should leave us with the new probability of our proposition p, taken by itself. This will be the probability of *p on the condition that* e *actually obtains*; hence, the conditional probability of p on e. Thus

$$NewPr(p) = \frac{oldPr(p \wedge e)}{oldPr(e)}$$

This is merely a simplification of Bayes' theorem. Let us look at that theorem, with the seemingly irrelevant occurrences of Ψ suppressed. We have

$$Pr(p_i/e) = \sum_{i=1}^{n} Pr(e/p_i) * Pr(p_i)$$

The denominator of this is equivalent to $Pr(e)$ by Theorem 13; so we also have it that

$$Pr(p_i/e) = \frac{Pr(p_i) * Pr(e/p_i)}{Pr(e)}$$

The numerator, by Axiom 2, is equivalent to $Pr(\mathrm{p}_i \wedge e)$, so we have

$$Pr(p_i/e) = \frac{Pr(p_i \wedge e)}{Pr(e)}$$

This allows us to explain an important phenomenon in probability — that of the reformation of our probabilities given our knowledge of the actual occurrence of a piece of evidence that was hitherto granted only to a probability. In fact, what this is, is a formula for calculating the effect of the changing of our probability of **e** from some number less than one to a probability of one. As such, it is a special case of a more general operation, namely, that of calculating the change in the probability values of one's propositions p_i given any change whatever in the probability of any of the evidence **e**. Such a function may be called a rule of *generalized conditionalization*.

In generalized conditionalization what we want to know is the probability of our arbitrary proposition p_1 given that the probability of our evidence **e** has now been changed to some new value (as opposed to its now being certain). If we can express this relationship in **ProbCal**, it has passed another test. Supposing that we can have a certain "base" of possible evidential states K_1, ..., K_n, each formed as a conjunction of the elements in some subset of our set of possible evidence e_1, ..., e_m, we may calculate the new probability of p_i by taking the sum of all

$$newPr(p_i) = \sum_{i} \frac{newPr(k_j) * oldPr(p_i \wedge k_j)}{oldPr(k_j)}$$

This gives us a formula for showing how all of the probabilities in a set of propositions are interconnected, and how we may obtain a method of calculating the effect a change in the probability of any one proposition in the system has on any other proposition in the system. This reveals an interconnection between all of the sentences of a given system.

This latter point is centrally important, as it shows that any system which works like **ProbCal** will display an integral interrelatedness between whatever works to fulfil the role of propositions. Of course we want these things which fill the role of propositions to be the beliefs that we have. What this is meant to show is that, if we try to pull one of these conservative manoeuvres, we shall be creating a situation in which we're allotting to some propositions a degree of belief which is much higher than it ought to be; a move that we know from our discussion of books will lead us to expect more "value" from the situation than it actually has. So in trying to be pleasantly surprised by returns higher than we "expected", we are putting ourselves at risk of being wrong more often than we would expect with regard to the converse of the proposition in question. The lack of isolation of the propositions/beliefs in our system forces us to make a "bet" on both sides of the proposition, rather than just one. Thus, the more conservative we are in one estimate, the more radical we are with respect to the mirror belief.

To a certain extent, then, this seems to be a particularly good description of what goes on in our belief systems. We can use this to represent formally how it seems that our beliefs do in fact interrelate. We can show why either optimism or pessimism are ultimately untenable positions, and even provide a sketch of the means by which we should arrive at a probability for a given proposition.

A valuable trait in any candidate for an inductive logic (in the narrow sense) is that it answer to certain intuitions we have about the interlocking nature of the concepts of "theory" and "probabilistic consequence". We seem to have a natural intuition of what it is to be a stronger or weaker consequence of a stronger or weaker theory. For example, we might have what seems to be a probable theory about some stellar phenomenon, and we might also have a second, wilder, improbable theory about the same phenomenon. Further, if it is right to say that the consequences of these theories can be expressed in probabilistic terms, then some consequences which are certain (upon the condition of the given theory's actually obtaining) would be assigned a value of 1 (relative to the given theory), those which are impossible according to that theory would be assigned a value of 0, and the rest of the spectrum of probability would be strung along the real interval $[0,1]$.

Granted the distinction between a weak consequence of a weak theory and a weak consequence of a strong theory, the probability of the weak consequence of the strong theory seems higher than that of a weak theory. What is more, the strength of a theory may be "diluted" in its consequences should those consequences be weak, while the strength of a consequence of a theory seems to be "diluted" as well by any weakness of the theory. It is necessary, then, that these types of intuitive relationships should be capable of being expressed by any decent candidate for an inductive logic.

ProbCal, while not itself a good candidate for such an overarching logic, does seem able to express these relationships. Let us examine its representation of the following:

(1) a strong consequence of a strong theory

(2) a strong consequence of a weak theory

(3) a weak consequence of a strong theory

(4) a weak consequence of a weak theory

1. Let $Pr(T_1/X) = .75$; this denotes a *high* probability of our theory's obtaining (given X). Moreover let $Pr(C_1/T_1 \wedge X) = .75$; this denotes a *high* probability that C_1 obtains, given that our theory T_1 obtains. Supposing (for the moment) that this is the only theory we have which allows for the possibility of the occurrence of C_1, we may use Theorem 13 to calculate the probability that C_1 obtains by using

$$Pr(C_1/X) = Pr(T_1/X) * Pr(C_1/T1 \wedge X) = .75 * .75 = .5625$$

2. Now suppose we have $Pr(T_2/X) = .25$; this denotes a *low* probability of our theory's obtaining (thus corresponding to a weak theory). Borrowing the value of a "strong consequence" from (1), we see that the probability of a strong consequence of a weak theory might be formulated as

$$Pr(C_2/X) = Pr(T_2/X) * Pr(C_2/T_2 \wedge X) = .25 * .75 = .1875$$

3. We give an example of a weak consequence of a theory T_1 by allowing that $Pr(C_3/T_1 \wedge \mathrm{X}) = .25$, a *low* probability. Thus we have it that we might represent a weak consequence of a strong theory by:

$$Pr(C_3/X) = Pr(T_3/X) * Pr(C_3/T_3 \wedge X) = .75 * .25 = .1875$$

4. Accordingly, with our usual numerical values for strength and weakness, the probability of C_4, should it be a weak consequence of a weak theory, will be:

$$Pr(C_4/X) = Pr(T_4/X) * Pr(C_4/T_4 \wedge X) = .25 * .25 = .0625$$

Constructing these probabilities via Theorem 13, we may see that the results do indeed appear do comport with out intuitions. C_1 is the most probable of the four consequences, as it was a probable consequence of a

probable theory, C_4 is the least probable as it is a weak consequence of a weak theory, and both C_3 and C_4 had similar effects of "dilution" quite in keeping with how intuitively we think such combinations work

This does not show, however, that **ProbCal** is sufficient to form an inductive logic. It merely shows that the axioms of **ProbCal** determine a system sufficiently rich to *express* these types of competing interests.

The problem remains that this system is too demanding with regard to the material at its disposal. That is, the interconnection proposed by the doctrine of conditionalization requires far more computation that we would actually expect any rational agent to do. This is interesting, as it suggests that we have a different overall understanding of what rationality is than is given by **ProbCal**. For example, one who works in strict accordance to **ProbCal** would be forced to make infinitesimal changes to any beliefs they had about stellar objects if they were forced to revise their belief about, say, their car's suspension. The only things which would be probabilistically isolated from the rest of their beliefs would be logical truths and falsehoods (and perhaps not even these). If someone actually did this on a regular basis, he would become so dysfunctional that we would quickly call him anything *but* rational. Real rationality requires us to make relevance judgements which actually violate the requirements of generalized conditionalization, but we do so without making book against ourselves. How could this be so, and where is the line that separates rational conformity to **ProbCal**, and irrational conformity?

So, again, we seem to be heading in the right direction with the formulation of the principles of **ProbCal**, but we cannot use it as the ultimate arbiter of rationality that we might have hoped it to be.

8 INDUCTIVE LOGIC IN THE NARROW SENSE

We have seen that **ProbCal** is not an inductive logic. That is, it cannot in the general case determine when, or to what extent, an argument is inductively strong. We have also see that, though not uncontested, there is a large body of opinion to the effect that if a logic of induction is ever to be constructed, it is best to construct it from **ProbCal** as a base. So regarded, an inductive logic will be an inductive logic in the narrow sense, that is, one which supplements **ProbCal**. Alternatively, we could regard our inductive logic as a non-conservative extension of **ProbCal**. So taken, there would be an inductive logic of middling breadth to add to our taxonomy of inductive logics. Thus inductive logic in the broad sense would be our quadruple $\langle$DefIndStr, **ProbCal**, Inductive Logic^N, a philosophical answer to Hume's Problem$\rangle$. Of middling breadth would be Inductive Logic^M, which is, **ProbCal** and Inductive Logic^N taken together. Least broad would be Inductive Logic^N taken by itself. It is clear that in as much as Inductive

Logic^N must conform to **ProbCal**, it is an arbitrary matter whether we opt for Inductive Logic^M or Inductive Logic^N. In what follows, "inductive logic" is taken either way unless otherwise indicated.

The inductive logic presented here, with certain inessential variations, is that of Carnap [1962]. It is the system **LFP**, after the title of Carnap's major work on the subject, *The Logical Foundations of Probability.*[11] In section 2.3 above, we said that on the logical interpretation, probabilities are degrees of rational belief. With Carnap, logical conception of probability takes an interesting twist. In **LFP** probability is likened to an objective quasi-logical relation that could be called *partial entailment.* In due course we shall note Carnap's efforts to reconcile the idea of degree of rational belief with the idea of degree of logical entailment.

We now specify for **LFP** a domain D of individuals and basic properties of individuals. **LFP** has proper names $a_1, \ldots, a_n$, a, b, c, ... for each object in D and one-place predicates (or predicate letters) $F_1, \ldots, F_n$, F, G, H, I, ... for each property in D. It will simplify the exposition to suppose that D contains exactly two objects, a and b, and exactly two properties F and G.

A predicate literal (alternatively a property literal) is any $\pm\Phi$ where Φ is a predicate letter $+\Phi$ is Φ and -Φ is $\ulcorner\neg\Phi\urcorner$. An open basic inventory on D is a (consistent) conjunction of literals such that each predicate letter Φ, has a single occurrence. If the language contains n predicate letters, then it possesses 2^n open basic inventories. Given our present assumption, there are just four open basic inventories of D.

$$Fx \wedge Gx$$

$$Fx \wedge \neg Gx$$

$$\neg Fx \wedge Gx$$

$$\neg Fx \wedge \neg Gx$$

For each object z, $z \,\varepsilon D$, there is exactly one basic inventory. Assuming every $z \,\varepsilon D$ to have one and only one formal proper name, we shall use the term "closed inventory" for the formal object obtained by substitutions of the formal proper name for z for the variable 'x' in the open basic inventory that describes z's condition.

We now specify Carnap's conception of a *state description* of the language of **LFP**.

> Let D be a domain with $a_1, \ldots, a_j \ldots$ as members. Then for each a_i, there is a basic inventory. The conjunction of these basic inventories is a state description of the language of **LFP** or, more briefly, an **LFP** state description or a state description on D.

[11]In what follows in this and the subsequent section, we follow Gustason [1994].

If D contains j objects and n basic atomic properties, then there are $2^j n$ state descriptions on D. In our present example, the following are examples of state descriptions on D.

$$Fa \wedge Ga \wedge Fb \wedge \neg Gb$$

$$Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb$$

It is well to notice that if K is a state description on D, then for each object and each property in D, K determines whether that object has that property or not. If we impose on the conjuncts of any K the requirement of pairwise semantic inertia, then we might think of any K as the specifier of a *possible world* on D.[12] On present assumptions, there are sixteen possible worlds on our two-object two-property domain. We suppose that one of these is the actual world on D.

The totality of state descriptions on D constitutes a set of mutually exclusive and jointly exhaustive alternatives, each true or false, as a matter of fact.

It is also supposed that state descriptions are thoroughly extensional. That is, extensional equivalence is a complete identity condition on state descriptions.

> **DefIndLog**^{N}: Let D be a domain and L a language for describing it. Then an inductive logic^N on $\langle D,L\rangle$ is the set **IR** for the real-valued function $Pr(\Phi/\Psi) = x$, for each pair of L sentences Φ and Ψ.

The rules of **IR** are different from the rules of **ProbCal**. Unlike the latter, the **IR** rules enable us to calculate the truth value of every statement of inductive probability which L is capable of formulating.

ProbCal gives us a way of characterizing unconditional probabilities. Φ's unconditional probability is Φ's conditional probability on a logical truth. Then, given Axiom 2 and ordinary arithmetic, it is provable that

THEOREM 15. $\Pr(\Phi/\Psi) = \dfrac{Pr(\Phi \wedge \Psi)}{Pr(\Psi)}$

Also provable is:

THEOREM 16. *Let υ be any real-valued assignment between 0 and 1 to the state descriptions of L on D and let the values of υ sum to 1. Then υ is an assignment of unconditional probability to the state descriptions in question.*

---

[12]Readers might recall that statements Φ, Ψ satisfy the principle *Logical Inertia* just when Φ neither implies nor is implied by Ψ and Φ and Ψ are jointly consistent. See chapter 2 above.

Proof. By Theorem 8 the probability values assigned over a set of mutually exclusive and jointly exhaustive sentences sum to 1. But the set of state descriptions of L on D is mutually exclusive on any condition and jointly exhaustive on any condition; so these probability values are unconditional. ■

Whereupon,

THEOREM 17. *Let Φ be any sentence of a language L. Then* $\Pr(\Phi)$*, its unconditional probability is determinable from the unconditional probabilities of L's state descriptions on D. Furthermore, the conditional probability* $\Pr(\Phi/\Psi)$ *is determinable from the unconditional probabilities of Φ and Ψ.*

Proof. By Axiom 1, Theorems 5 and 15. ■

Jointly Theorems 15 and 16 enable us to have:

THEOREM 18. *Let υ be an assignment of values to the state descriptions of L on D, and suppose that the values sum to 1. Then υ plus* **ProbCal** *determines an inductive logic for D.*

To see how Theorem 18 works, and also to expand upon the proof of Theorem 17, let us consider an example as it applies to $\langle D, L\rangle$. Suppose that we wish to determine the conditional probability, $Pr(Gb/Fa \wedge Ga \wedge Fb)$. We know by deductive logic that $\ulcorner\Phi \equiv (\Phi \wedge \Psi) \vee (\Phi \wedge \neg\Psi)\urcorner$. Hence, "$Fa \wedge Ga \wedge Fb$" is logically equivalent to "$(Fa \wedge Ga \wedge Fb \wedge Gb) \vee (Fa \wedge Ga \wedge Fb \wedge \neg Gb)$".

In other words, "$Fa \wedge Ga \wedge Fb$" is equivalent to any state description on $\langle D, L\rangle$ in which it occurs.

Theorem 5 and Axiom 1 now provide that

$$\begin{aligned} & Pr(Fa \wedge Ga \wedge Fb) \\ &= Pr(Fa \wedge Ga \wedge Fb \wedge Gb) \vee (Fa \wedge Ga \wedge Fb \wedge \neg Gb)) \\ &= Pr(Fa \wedge Ga \wedge Fb \wedge Gb) + Pr(Fa \wedge Ga \wedge Fb \wedge \neg Gb) \end{aligned}$$

If we put it that the probability of the state description "$Fa \wedge Ga \wedge Fb \wedge \neg Gb$" is 1/10, and that the probability of the state description "$Fa \wedge Ga \wedge Fb \wedge \neg Gb$" is 1/20, then $Pr(Fa \wedge Ga \wedge Fb) = 1/10 + 1/20 = 3/20$. Applying Theorem 15 we see that

$$Pr(Gb/Fa \wedge Ga \wedge Fb) = \frac{Pr(Fa \wedge Ga \wedge Fb \wedge Gb)}{Pr(Fa \wedge Ga \wedge Fb)} = \frac{\frac{1}{10}}{\frac{3}{20}} = \frac{2}{3}$$

We now define the reflexive and the symmetrical relation of isomorphism.

> **DefIso**: If S and S^* are state descriptions, then S is *isomorphic* to S^* iff S (or S^*) can be transformed into S^* (or S) by repeated interchange of proper names, each occurrence of a given name replacing all occurrences of some.

For example, "$\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb$" and "$Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$" are isomorphic to each other. For the result of exchanging proper names in the first is "$\neg Fb \wedge Gb \wedge Fa \wedge \neg Ga$", which is equivalent to the state description "$Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$". Intuitively, it may be said that these isomorphic state descriptions are descriptions of *qualitatively similar* possible worlds, in each of these an object has G and lacks F and an other object has F and lacks G.

We may now say that

> **DefIsoMes**: The *isomorphic measure*, $i(S)$, of a state description S is the number of state descriptions isomorphic to S (including S, since isomorphism is reflexive).

Hence $i(\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb) = 2$. This state description is isomorphic to itself and also to "$Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$". Moreover, any two isomorphic state descriptions have the same isomorphic measure.

A final definition:

> **DefStrucDes**: A *structure description* is a disjunction of all state descriptions isomorphic to one another.

In the limiting case in which an S is isomorphic to itself alone, S it its own structure description.

In the table below we list all sixteen Ss, or state descriptions, on $\langle D, L\rangle$. Those that are isomorphic to one another are enclosed by braces. In each case, the embraced items are disjuncts of a given structure description.

State Description (S)	**Isomorphic Measure i(S)**	**Pr(S) in LFP**
1. $Fa \wedge Ga \wedge Fb \wedge Gb$	1	2/20
2. $Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb$	1	2/20
3. $\neg Fa \wedge Ga \wedge \neg Fb \wedge Gb$	1	2/20
4. $\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb$	1	2/20
5. $Fa \wedge Ga \wedge Fb \wedge \neg Gb$	2	1/20
6. $Fa \wedge \neg Ga \wedge Fb \wedge Gb$	2	1/20
7. $Fa \wedge Ga \wedge \neg Fb \wedge Gb$	2	1/20
8. $\neg Fa \wedge Ga \wedge Fb \wedge Gb$	2	1/20
9. $Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb$	2	1/20
10. $\neg Fa \wedge \neg Ga \wedge Fb \wedge Gb$	2	1/20
11. $Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$	2	1/20
12. $\neg Fa \wedge Ga \wedge Fb \wedge \neg Gb$	2	1/20
13. $Fa \wedge \neg Ga \wedge \neg Fb \wedge \neg Gb$	2	1/20
14. $\neg Fa \wedge \neg Ga \wedge Fb \wedge \neg Gb$	2	1/20
15. $\neg Fa \wedge Ga \wedge \neg Fb \wedge \neg Gb$	2	1/20
16. $\neg Fa \wedge \neg Ga \wedge \neg Fb \wedge Gb$	2	1/20
		Sum: 1

Inductive logic in the narrow sense can now be specified. $\mathbf{IndLog}^N = \mathbf{LFP}$. $\mathbf{LFP} = \langle Axiom1, Axiom2, Axiom3, Axiom4, Axiom5, Axiom6\rangle$. That is, $\mathbf{LFP}^N$ is defined by the five **ProbCal** axioms, to which a further axiom, the Axiom of Induction, is added. This latter is

AXIOM 6. For each state description S, $Pr(S) = 1/(i(S)*n)$, where n is the number of structure descriptions.

In our table, there are ten structure descriptions. Axiom 6, which is compatible with Theorem 8, fixes an assignment of unconditional probabilities to all state descriptions on $\langle D, L\rangle$. Each is given a value, $Pr(S)$, which is inversely proportional to $i(S)$, and these values sum to 1, as can be seen from the table.

How plausible is our axiom, Axiom 6? In the interests of space, we illustrate how Axiom 6 works with reference to the rule of induction by simple enumeration, perhaps the most basic inductive rule of all. Let F and G be any properties. The inductive probability that the next individual observed to be an F will also be a G increases as the number of already observed Fs that are G also increases, provided that there are no negative instances and that 'Fx' and 'Gx' are logically independent. The addition of **AxInd** to **ProbCal** is meant to give probabilities for expressions in the form $Pr(\Phi/\Psi)$ that conform to what we regard as standard inductive criteria, indeed, to rules such as induction by simple enumeration. Thus, as applied to our simple D, the rule operates as follows.

"$Fb \wedge Gb$" is logically equivalent to the disjunction of state descriptions recorded in our table as (1), (6), (8) and (10). (It occurs in each and in none other). Similarly, "Fb" is logically equivalent to the disjunction of those same state descriptions further disjoined with (2), (5) , (12) and (14). Accordingly,

$$Pr(Gb/Fb) = \frac{Pr(Fb \wedge Gb)}{Pr(Fb)} = \frac{\frac{1}{10} + 3(\frac{1}{20})}{2(\frac{1}{10}) + 6(\frac{1}{20})} = \frac{1}{2} = 1/2$$

The probability that b is black given that it is a raven is 1/2. By the rule of induction by simple enumeration, the probability will increase as the number of uncontradicted positive instances increases. So we have it that

$$Pr(\mathrm{Gb/Fa{\wedge}Ga{\wedge}Fb}) > Pr(\mathrm{Gb/Fb}).$$

Standard inductive practice assumes that the world is uniform. More carefully, the more uniform a domain of application, the more reliable are the rules of standard inductive practice. This thought is enshrined in the conditions that define **LFP**. We can see how by observing that a state description's isomorphism measure *varies inversely* with the degree of uniformity exhibited by the "possible world" it specifies. That is, the larger the $i(S)$ the less uniform the world specified by S. In our table, **LFP** gives higher unconditional probabilities to worlds exhibiting greater uniformity.

We may now re-pose Hume's question. Can we give any satisfactory, noncircular justification for giving higher unconditional probabilities to Ss specifying uniform worlds? If it is replied that such assignments give rise to a logic that has worked well in the past this places us squarely on Hume's carousel, and if this is the only plausible answer, it is blatantly circular, as Hume saw.

Carnap himself is one theorist of a group that holds that the foregoing answer is not the only candidate with which to meet Hume's challenge. On this view, inductive probability is a purely logical relation and hence statements on inductive probability, where true, are logically true and, where false, are logically false. Such theorists understand "logical" in a somewhat relaxed way. What they intend is that determining the truth (or falsity) of a statement of inductive probability on statements Φ and Ψ is simply a matter only the general techniques of logic together with the meanings of Φ and Ψ. More secure is the claim that statements $Pr(\Phi/\Psi) = n$ are not *empirical* statements. In this they bear a closer similarity to mathematical propositions than to logical tautologies.

We see in this the beginnings of a possible answer to Hume. Statements of inductive probability, like statements of mathematics, are, if true, non-empirically true. Further, these truths conform to standard inductive practice, many principles of which, including some of the most sophisticated, are theorems of **LFP**. Such principles are justified to the same general extent that mathematical and set theoretic principles are justified. This is not the whole justification of induction. Another substantial claim lies in wait. It is that the truths of inductive probability constitute a standard of rationality under conditions of uncertainty. So, if it is a truth of **LFP** that $Pr(\Phi/\Psi) = n$, then n is the degree to which a cognitive agent is justified in believing Φ on evidence Ψ. More precisely, if agent **A** has verified that $Pr(\Phi/\Psi) = n$, then n is a fair betting quotient for **A** in relation to a bet on Φ given Ψ, provided the stakes are small enough in relation to **A**'s total wealth. Further, n is the factor by which the value of the events or facts described by Φ must be multiplied to generate for **A** a rational expected utility of the course of action he has under consideration. If **A** is rational he will believe a (real) logical truth to degree 1, and that a club will be drawn from a standard deck to degree 1/4; and he will accept a bet on this latter only at odds 3:1 or higher. Now if this is right, the theorist may wish to deal with Hume by saying that since true statements of inductive probability are normatively canonical for agents acting in conditions of uncertainty, the justification of induction inheres in the non-empirical theorems that constitute such canonicity. We shall return to this point.

The view under consideration originated with Keynes [1921] and was considerably refined in Carnap [1949] and [1952].

9 THE LOGICAL THEORY OF PROBABILITY

We see that according to **LFP**, inductive probability is a strictly non-empirical relation on pairs of sentences. Probability theorists have accustomed themselves to speak of the relation as strictly logical. But, as already noted, it is better to think of **LFP** as a logical theory, if at all, in much the same loose way that mathematicians sometimes think of the abstract theory of pure sets as a logical theory.

Of greater importance is to be clear about what the logical (probability) theorists, so called, intend to convey by characterizing their central target notion as purely logical. What is intended is that probability is an objective relation which holds on proposition-pairs, a relation that obtains or does not, independently of what cognitive agents or belief-revisers may think. In other words, probability is regarded as a propositional relation whose career is unaffected by whether or not any believers ever actually exist.

Of course, actual believers do in fact exist in considerable abundance. This being so, there exist the beliefs of those believers and there exists the plain fact that those beliefs are held with various degrees of confidence. These are subjective matters.

Carnap, and others, hold that an agent's belief-revisions are rational to the extent that they conform to the relevant theorems of **LFP**. What hasn't yet been made clear is what *conformity* here means. In other words, we haven't yet said what are the detailed conditions that must be met before an agent's doxastic behaviour can be said to conform to the requirements of **LFP**. Let us see.

We begin with A's subjective degree of belief with respect to a proposition Φ at a time t. We shall say with Carnap that the subjective degree of A's belief at t that Φ is the highest betting quotient under which A is prepared to bet on Φ at t, where the stakes are realistically small. Now it is routinely the case that A's disposition to lay such a bet will depend on the state of A's relevant knowledge, including her observational knowledge. Suppose then that Ψ is A's observational knowledge at t, and let us restrict ourselves to Ψ's influence on A's state of belief at t with regard to Φ.

Then the credence of Φ, given Ψ or, $Cr(\Phi/\Psi)$, is the degree to which an **A** would believe Φ were Ψ her total evidence. It is noticeable at once that this is a conception of credence that gives ample recognition to the fact that different people assign different values to (Φ, given Ψ), and in that respect Cr is a realistic notion. On the other hand, it should also be remarked that in this account there is no explicit deployment of temporal parameters. This alone shows that Cr is not a realistic credence function in all respects. What Cr is intended to capture is A's overall disposition to revise beliefs on the basis of observation. Hence Cr is a more stripped down mechanism than any agent's real-time credence functions. It might even be thought of as a meta-credence function, as a kind of higher order mechanism that

coordinates the operation of a whole host of real time credence-operators.

Even so, it is clear that different people can exhibit different *Cr*s in similar circumstances. It is also clear that *Cr*-functions need not be rational. This being so, what seems to be needed is the specification of a *Cr*-function whose use would give rise to rational behaviour under conditions of uncertainty. Or, as Carnap would say, inductive probability is a rational *Cr*-function. There is some question as to whether there is just one such function. Among logical theorists, the consensus is that there are many. Reasonable agents might, in nearly identical circumstances, display different *Cr*-functions.[13] Abstract though it certainly is, *Cr* is also a plainly psychological function. This leaves us to wonder what role *Cr* has to play in any inductive logic in which inductive probability is a logical (i.e., non-empirical) relation.

In a number of theories of induction, probability is equated with confirmation, each being binary relations on propositions governed by non-subjective or non-psychological conditions. Although confirmation, like belief, comes in degrees, the logical theorist will always hold that the confirmation relation of inductive logic will take on its values independently of values that may accrue to an agent's *Cr*-function. Thus if we express confirmation claims in the canonical notation

$$c(\Phi/\Psi)$$

which expresses the degree to which Φ is confirmed on Ψ, then the point at hand is that for appropriate n and m

$$c(\Phi/\Psi) = n$$

is true or false independently of whether

$$Cr(\Phi/\Psi) = m$$

is true or false. Since *c*-values are determinable entirely objectively and *Cr*-values not, it may seem to follow that *c* and *Cr* are wholly different things. This is not how the logical theorist sees it. Just as a syllogism can be used as a refutation of an opponent's thesis, so too can a *c*-function serve as an actual agent's *Cr*-function. It is in principle possible that there are several different *c*-functions just as there are several different *Cr*-functions. Of the totality of *c*-functions, we are interested in those that are *inductively adequate*. To this end, it is proposed that

> **IndAdCon**: A *c*-function is inductively adequate to the extent to which if used by **A** as his *Cr*-function it would induce in **A** rational behaviour.

[13]Thus Carnap changed his mind from one rational *Cr*-function (Carnap [1949]) to several (Carnap [1952]).

So long as it can be supposed that the conditions of **IndAdCon** are fulfillable without resort to any psychological factors, then our inductive logic seems to escape the charge of psychologism; and we see in this Carnap's attempt to unify the rational belief and partial entailment conceptions of logical probability.

LFP can be as defined by a particular c-function which Carnap denotes as c^*. Thus $c^*(\Phi/\Psi)$ is the value assigned by probability functor if **LFP** to $\ulcorner\Phi$, given $\Psi\urcorner$. To specify this function we make an assignment of values, summing to 1, to the state descriptions of an L for a D. **LFP** is then used to specify the unconditional value of any statement expressible in L and the conditional value of any such statement given any other such statement. In particular, $c{*}(\Phi/\Psi)$ is determinable via the theorems of **ProbCal** applied to the probabilities assigned to state descriptions in accordance with Axiom 6.

There are infinitely many c-functions, many of which closely resemble c^*. Thus there are infinitely many inductive logics, many of which resemble **LFP**. Carnap holds that only c^*, and those that resemble it, possess sufficient inductive adequacy to qualify for use as a rational agent's Cr-function. The c^* determines higher unconditional probabilities for state descriptions describing more uniform possible worlds. It is a distinctive feature of Carnap's work on inductive logic that there exists a method for distributing all such inductive logics over a continuum, where a logic's position on the continuum represents the extent to which it facilitates learning from experience. Learning from experience comes to this: Take some given formula $c(Gb/Fb)$ and a "later" formula $c(Gb/Fa \wedge Ga \wedge Fb)$. If from the one to the other there is a large increase in c-value, there has occurred a correspondingly substantial amount of learning from experience. Thus learning from experience for some given c-function is a matter of how c shapes our expectations on the basis of our experience. Every c-function on Carnap's continuum works in such a way that the c-value that b is black given that it is a raven and that an inspected a is both black and a raven will be greater than the c-value that b is black given merely that it is a raven. Even so, these c-functions will differ by the extent to which they give different, although always greater, c-values for the cases in question.

Inductive probability is a rational Cr-function which, in turn, is a c-function having inductive adequacy which, further, is a c-function occurring on Carnap's continuum. What does this inductive adequacy consist in? Carnap's answer is that it consists of satisfaction of the following axioms:

> Axioms 1–5 of **ProbCal**, in which 'c' replaces all occurrences of 'Pr'.

AXIOM 7. For any $\langle D, L\rangle$ giving at most a finite number of possible worlds, if $c(\Phi/\Psi) = 1$, then $\models \Psi \supset \Phi$.

Note that Axiom 7 represents a decision to preclude the making of a partial

Book against any reasoner modelling the c-function in question. (Recall that, Axiom 7 is the converse of Axiom 3.)

A c-function is *regular* if and only if its assignments of values to state descriptions on $\langle L, D\rangle$ sum to 1 and comply with Axiom 1 to Axiom 7, provided L is a finite language. It is easy to see that C^* is regular. The further axioms, 8 to 13, are the axioms of *invariance*.

AXIOM 8. The value of $c(\Phi/\Psi)$ does not change under any finite permutation of D-objects.

AXIOM 9. The value of a $c(\Phi/\Psi)$ does not change under any permutation of D-predicates of any family.

'Family' is new to our exposition. A family is a set of predicates denoting mutually exclusive and jointly exhaustive properties. In the simple model with which we are working, D's total set of predicates is the sole family of such predicates, which need not be mutually exclusive and jointly exhaustive.

AXIOM 10. The value of $c(\Phi/\Psi)$ doesn't change under any permutation of predicate-families of the same cardinality.

AXIOM 11. Provided that Φ and Ψ are truth functional sentences, the value of $c(\Phi/\Psi)$ is unchanged under any extension of D's objects.

AXIOM 12. The value of $c(\Phi/\Psi)$ doesn't change under arbitrary supplementation of D's predicate families.

AXIOM 13. Let freq (π) be the relative frequency with which property π has shown up in n trials, (that is, $\text{freq}_n(\pi) = \frac{s}{n}$ is true in that state of affairs in which s out of n times, involving objects $a_1, \ldots, a_n$, π has been observed.)

Then,

$$c(\pi a_{n+2}/\text{freq}_n(\pi) = \text{s/n} \wedge \pi a_{n+1}) > c(\pi a_i/\text{freq}_n(\pi) = \tfrac{s}{n})$$

Intuitively, the degree of confirmation that an object has a certain property, given both that a second object also has it and that the relative frequency with which objects having that property is s in n observations, is greater than the degree of confirmation that some object has that same property given only the same frequency of its occurrence.

Axiom 13 is of central importance. Without it, it would be next to impossible to recast what we have been calling the principles of our standard inductive practice as theorems of any inductive logic of the **LFP** sort. Axiom 13 achieves this centrality by setting forth the basic mechanics of what Carnap understands by learning from experience.

The function c^* obeys all our axioms. If Carnap is right, c^* is inductively adequate and when used by a reasoner constitutes a rational Cr-function for him. But what justifies the assertion that satisfaction of these axioms

produces inductive adequacy? Indeed, what reason is there even for accepting these axioms? It is useful to consider whether these questions might be answerable in a way that also constitutes a satisfactory reply to the challenge of Hume's Problem. We may take it as given, as Carnap himself does, that if the axioms were justified as well-confirmed empirical conclusions of inductive arguments, their justification could not in turn count as a satisfactory answer to Hume. On the other hand, if the axioms are justified non-empirically and in an *a priori* way, for example, in the manner in which mathematical axioms are said to be justified, then it would be open to construct the following answer to Hume:

> **AnsHum**: Axioms 1–13 are *a priori* principles and they imply the principles that underlie our standard inductive practices.

The answer is unsatisfactory. It is true that if the axioms are justified in the same way that *true* mathematical axioms are justified, and if it is also true that such justification shows what it justifies to *be* true, then it would follow that the Axioms 1–13 are true *a priori*. It would follow that the rules of induction, being consequences of *a priori* truths, are also true. Jointly, it might be said that Axioms 1–13 constitute an *a priori* justification of induction. This, certainly, is what Carnap himself did think.

Carnap was wrong. What he showed was that if the axioms are justified in a way that shows them to be true, then, since the principles of induction are theorems of those axioms, there would exist a justification of those principles which shows them to be true. What was not shown is that the axioms are justified in ways that shows them to be true. Repairs were sought by a number of theorists, notably John Kemeny (Kemeny [1963]). They came by way of a specification of *adequacy conditions*. We want to know when a c-function serves adequately as a rational Cr-function. One possibility is that a c-function works properly as a rational Cr-function when the axioms on the c-function enjoy substantial *intuitive* support. We note, in passing, that any such appeal to intuitiveness is at bottom an appeal to what "we" think are good inductions. What we think are good inductions are embodied in standard inductive practice. Thus inductive axioms and principles are, on this view, those that harmonize with standard practice. So an appeal to adequacy conditions is tantamount to a justification by reflective equilibrium, of the sort made famous by Goodman (though not under that name).[14]

Kemeny proposes four conditions of adequacy, each in his opinion fulfilled by Carnap's logic.

> **AC1**. No c-function is adequate if it specifies a wagering-system that permits the making of Dutch or partial Books.

[14]The baptism may be found in Rawls [1971].

If this is right, then **AC1** would seem to be satisfied on account of Axioms 1–7.

> **AC2**. The value of $c(\Phi/\Psi)$ should not be fixed or partially fixed by extraneous considerations, i.e., by propositions other than Φ and Ψ. If this is right, Axiom 11 and Axiom 12 would seem to guarantee its satisfaction.
>
> **AC3**. Proper names of objects should be semantically inert with respect to c-values. That is, $c(\Phi\alpha) = c(\Phi\beta)$, for any proper names α, β.

If **AC3** is correct, it would seem to support Axioms 8–10.

> **AC4**. c-functions must be constructed in such a way as to model the phenomenon known as "learning from experience".

Clearly, **AC4** purports to justify Axiom 13.

Carnap's c^*-logic is beset with a number of technical problems for whose discussion we lack space. We note in passing that in infinite domains, the c-value of generalizations in the form $\ulcorner\forall v(\Phi v \to \Psi v)\urcorner$ is zero to begin with, and it stays zero no matter how impressive the confirming evidence. Even in large finite domains, such generalizations begin at close to zero and remain there even in the face of lots of confirming instances. Technical extensions have been devised for Carnap's basic system in the hope of accommodating the confirmation of generalizations. (See especially Hintikka [1966]).

Arguably the greatest setback to Carnap's programme for inductive logic comes from a powerful and clever argument of Hilary Putnam (Putnam [1975], [1975b]). If correct, it puts the Carnapian programme into permanent retirement. That alone would make it a significant development, but Putnam's attack is all the more significant because it seems to count against any Bayesian approach to induction.

10 CONTRA BAYESIANISM

Recall what is intended by the term "Bayesian". A Bayesian is someone who eschews "qualitative" approaches to confirmation. If you are a Bayesian, the only satisfactory way of accounting for how evidence weighs upon hypotheses and scientific theories is one in which degrees of belief are held to strict conformity with the **ProbCal** axioms. What is more, a Bayesian theory will always give a central place to Bayes' Theorem, and will attempt to justify the principles of probability by examining the relationship between degrees of belief and betting behavior. Thus a Bayesian approach is always a context for the development of Dutch Book arguments. A further, and related point on which Bayesians agree is that learning from experience should be

understood as conditionalization. Thus there is a *rule of strict conditionalization* which Bayesians espouse. By this rule, if it is learned with certainty that Φ then the probability functions Pr^{old} and Pr^{new}, representing in turn degrees of belief prior to and after the new information Φ is required, are related by the equation.

$$Pr^{new}\Psi{=}Pr^{old}(\Psi/\Phi)$$

Bayes' Theorem attempts to justify this claim by making it precise how the recognition of new evidence acts on previous belief strengths to yield new belief strengths. (A subtle extension of conditionalization in which allowance is made for learning with less than certainty has been developed in Jeffrey [1983]).

To return to Putnam, we concentrate on a hypotheses formulable in the language L of first order arithmetic, with names $a_1, \ldots, a_n, \ldots$ denoting concrete individuals, and empirical predicates with which to ascribe properties to those concrete individuals.

Consider a Bayesian agent A_1 who attaches degrees of belief to statements expressible in L. According to Putnam, it is desirable to hold A_1 to fulfillment of two conditions. One, call it D_1, is that A_1 should be capable of learning any true *and effective* hypothesis. The other, call it D_2, is that for every compound predicate Φ of L and any $n{\geq}0$ there is an $m{>}0$ such that

$$D_2 : Pr(\Psi\alpha_{n+m+1} / \sum_{i\leq n} \pm\Psi\alpha_i * \sum_{n<j\leq n+m} \Psi\alpha_j)\ 0.5$$

In other words, no matter which of the first n individuals have the property (expressed by Ψ) or not, if the next m objects all have Ψ for a sufficiently large m, then the probability that the next object α_{n+m+1} also has Ψ is greater than even odds or "chance". As it happens, this second desideratum is a consequence of the first.

Let us suppose that the predicate R expresses the property of redness, and that K is an infinite set of integers $n_1, n_2, \ldots$ such that the probability of $\alpha_{n_{i+1}}$ having R exceeds 0.5 if all preceding objects after α_{n_i} are also red. If D_2 is true, we can be sure that K exists.

Suppose next that $\mathbf{A}_1$'s probability function Pr is minimally recursive, that is, that Pr is effectively computable for truth-functional compounds of atomic empirical sentences. If this supposition is correct, K is a recursive set. Then, as Gödel showed, there will be a predicate Q exactly definable via '+' and '$\times$' of which K is the extension. (For details, see Rogers [1967, chapter 14]). This enables the formulation of the "diagonalization" hypothesis:

Diag: $\forall_i(Q_i \leftrightarrow \neg R\alpha_i)$

in which the variable of the quantifier ranges over the natural numbers.

Now *Diag* is effective and it could well be true, but its instantial confirmation can't reach and sustain a level greater than 0.5. Thus *Diag* violates **D2**. Consequently, for any theorist of Bayesian stripe whose probability function is minimally recursive there will be good hypotheses that can't be learned, even in the weak sense in which learning is just a matter of instantial confirmation. We illustrate the point as follows.

A positive instance, $\mathbf{I}(n)$, of the quantificational scheme *Diag* will be in the form $(Qn \wedge \neg R\alpha_n)$ or $(\neg Qn \wedge \mathrm{R}\alpha_n)$. Given that evidence statements are true and that if Qn is true (or false) in any model of L it is true (or false) in every model of L, we get the following result. For the positive instances of *Diag.* we can determine probabilities: $Pr(Qn \wedge \neg R\alpha_n) = Pr(\neg R\alpha_n)$, and $Pr(\neg Qn \wedge R\alpha_n) = Pr(R\alpha_n)$. From this,

$$Pr(Diag/\sum_{i\leq n} I(n)) = Pr(Diag/\sum_{i\leq n} I*(n))$$

where $I^*(n)$ is $\neg R\alpha_n$ or $R\alpha$ according as $I(n)$ is $(Qn \wedge \neg R\alpha_n)$ or $(\neg Qn \wedge R\alpha_n)$. Accordingly, evidence sequences will be in the form $R\alpha_1 \wedge R\alpha_2 \wedge \ldots$. The rest is intuitive upon reflection, as Arthur Burks often says.

11 NON-PASCALIAN PROBABILITY

The inductions in which we have been interested are those that one might call "ampliative". They involve the making of claims or the holding of beliefs that "go beyond" and "say more" than the evidence on which they are based.

The inductive logician wants to know the conditions under which ampliative inferences are sound. There are two possible answers to take account of. Given that it is positive instances which lend inductive support to generalizations which kind of positive instances is it better to have:

(a) those that exhibit significant variations (Bacon)?

(b) those that exist in large numbers (Carnap)?

If we favour answer (a) then we have a preference for "variative" or *eliminative* induction. If, on the other hand, we incline toward answer (b) we are (or tend to be) *enumerativists* about induction. A question for the inductive logician, therefore, is which conception of induction is the right one or the better one.

We have already emphasized the importance of getting clear about the connection between inductive support (or inductive strength) and the kind of probability that **ProbCal** faithfully describes. One of the interesting features of Mill's *A System of Logic* is the interest it shows in analyzing the

general concept of greater-or-less inductive support by way of real numbers in the interval $[0 - 1]$. Following Cohen, we shall speak of this generic conception of inductive support as *Baconian induction*, and of the tradition in probability theory ensuing from the work of Pascal and Fermat as the *Pascalian tradition* and, correlatively, *Pascalian probability*.

It is important that we keep in mind that Baconian induction and Pascalian probability each comes in somewhat different forms. Here is a rough taxonomy.

Baconian Induction	**Pascalian Probability**
$\Downarrow$	$\Downarrow$
Eliminative (Bacon, Mill)	Logical (Keynes, Carnap)
Enumerative (Carnap, Hintikka)	Frequency Orientated
	(Reichenbach, von Mises, Salmon)

In principle, any variation of Pascalian probability (each of which obeys Pascal's rules of chance, which are predecessors of the theorems of **ProbCal**) can be used to elucidate any variation of Baconian induction, i.e., the concept of more or less inductive support or inductive strength.

A third variation is truly a thing apart, *sui generis*. It too satisfies Pascal's rules of chance, and is indeed our friend Bayesianism. According to this view, probability elucidates degrees of belief, not grades of evidential support. If you are a Bayesian about probability you avoid the necessity of adjudicating the dispute between eliminative and enumerative induction. In other words, if you are a Bayesian about probability, in no straightforward sense can you use probability to elucidate Baconian induction, not anyhow unless you are prepared to argue that a suitable extension of **ProbCal** qualifies as a normative model of belief-revision and the normativity of the model consists in the correlation between:

> Belief-revisions R in the light of beliefs $B_1, \ldots, B_n$.
> The objectivity of generalization G's degree of support offered by evidence $E_1, \ldots, E_n$, where the content of R is G and of the B_i are the E_i.

It is high time that we marked some key distinctions. We have linked two key concepts: induction and probability. They are linked (so we suppose) in some significant way. But, as we have seen, there is no one thing that probability is even when the axioms of **ProbCal** are honoured, and there is no perfectly homogeneous concept of induction.

Probability is a case in point. It presents us with what we might call the problem of One and Many. Note well, we are here speaking of *Pascalian* probability. Pascalian probability satisfies the following conditions.

(a) it fulfills the axioms of **ProbCal**

(b) there are different kinds of Pascalian probability.

How are (a) and (b) jointly possible? The answer is that Pascalian probability is representationally (or mathematically) just one thing. Probabilities are (represented as) real numbers constrained by Axioms 1–5. Even so, given the clear differences between the logical, frequency and Bayesian approaches to probability, we must acknowledge probability's *ontological multiplicity* (or, if you like, its *conceptual* or *interpretational* multiplicity). Thus Pascalian probability is the same mathematical (or representational) thing no matter what it is ontologically (or conceptually or interpretatively). Carnap thinks that what probability really is is a quasi-logical relation on proposition-pairs. Reichenbach thinks that probability really is a frequency-based concept of a certain kind. And (descriptive) Bayesians think that what probability really is is a concept of degrees of belief. Now it is clear that nothing can be a quasi-logical relation and a frequency-based concept and a concept of degree, of belief. So, ontologically, probability is more than one thing. Probability is mathematically One and ontologically Many. All three conceptions of probability satisfy the **ProbCal** axioms. So we may say that the mathematics of probability underdetermines the ontology of probability.

We now see that whether probability is a good elucidator of induction depends upon what induction is (eliminative or enumerative) and what probability is *both mathematically and ontologically.*

It is well to note that Bacon bears upon our discussion in two ways. First the basic idea that generalizations are confirmed on the evidence of their positive instances or, somewhat more broadly, that the epistemic health of a generalization is a matter of what evidence there is for it, has given rise to the name "Baconian induction". But there is also a second, and narrower, concept of Baconian induction, namely eliminative induction. A generalization is eliminatively strong to the extent that its trial history is sufficiently varied and representative that the prospect of the experiment finding a counterexample is judged, however provisionally, to have been eliminated. Thus eliminative induction favours the finding of positive instances (and no negative instances) across the board of potentially falsifying situations. It is not just the number of positive instances that count, but also the exhaustiveness of their representation of different varieties of potentially falsifying conditions; hence, eliminative induction may also be called *variative.* (Cohen [1989, 37–38 and 145–163]).

Whether in the broad or narrow sense, Baconian inductions issue in conclusions or hypotheses of differing grades of certainty. It is one of Mill's achievements in the *Logic* to give mathematical expression to this idea of level, or degree, of certainty. This he did by appropriating **ProbCal** as the

elucidating device. In this he was almost certainly anticipated by the Port Royal logicians (Arnauld and Nicole [1996]) in a section of *Logic or the Art of Thinking* that may have been written by Pascal himself. It is crucial that we remember that **ProbCal**'s concept of probability arose from the study of the mathematical theory of games of chance, such as dice. The Latin for "die" is *alea* and for "dice-player" is *aleator*. Some theorists (e.g., Cohen [1989]) take this as occasion to call Pascalian probability "aleatory". Mill's achievement was the appropriation of aleatory probability in the analysis of induction.

Not forgetting its seventeenth century anticipation, by the time that Mill had produced his *Logic*, it was rather widely held by theorists that a statement's *degree of certainty* and a move in a game's *chances* share a common mathematical structure. So it is clear that at some time or other from the last third of the seventeenth century to the middle of the nineteenth a second notion of probability emerged. It is the probability that attaches to propositions to the extent to which they are well-evidenced, and may be spoken of as *inductive probability*.

Probability comes thus in two kinds, aleatory and inductive. Here, too, we come upon two things and one. Aleatory and inductive probability differ ontologically, but as Mill and a very substantial tradition which followed him are to be believed, they are the same mathematically.

Difficulties. As we have seen, in situations of chance, as when dice are being tossed or playing cards are being drawn, the conjunction axiom of **ProbCal** is unproblematically true for compounding probabilities. The basic idea, which the axiom honours, is that for statements Φ and Ψ the probability of $(\Phi \wedge \Psi)$ is, except for limiting cases, guaranteed to be less than the probability of Φ or of Ψ.

But consider this example from Cohen [1989, 19]. Suppose Christie's New York attests to the genuineness of two Vermeers which a gallery is considering for purchase. Given the track record of Christie's, the probability that picture one is a Vermeer is high, and the probability that the second picture is a Vermeer is also high. (Suppose that each picture receives an assessment to the point of moral certainty.) In such a situation, it is not true that the proposition that they are both genuine Verneers has to be less well attested to than the proposition that either is. So here is a case in which evidence for a proposition is inductive but not Pascalian. In other words, here is a case in which probabilistic induction in a non-aleatory context and probabilistic induction in an aleatory context do not coincide; they are not mathematically equivalent.

A second case. Not much is known about certain features of Julius Caesar. On the available evidence, there is not much reason to think that Caesar had a mole on the inside of his left forearm, placed three inches from the wrist's pulse point. But neither is there much reason to say that he didn't, since moles are fairly liberally distributed over human epidermi, and the

evidence in this case is scant. Neither proposition is even remotely credible on the evidence. In so saying, we run afoul of the **ProbCal** guarantee that $Pr(\Phi) + Pr(\neg\Phi) = 1$. That is, there are cases in which the sum of two utterly low probabilities, is itself rather low, even when the prior probabilities apply to propositions that are one another's negations.

A third case is one in which the probability that Spike "did it", given the evidence before the Court *seems* to equal the improbability of that evidence's being before the Court, given that Spike didn't do it after all. But if this is right we have a situation of the form

$$Pr(\Phi/\Psi) = Pr(\neg\Psi/\neg\Phi)$$

which is not authorized by **ProbCal**. However, if we could show it true — at least for certain cases — that $Pr(\Phi/\Psi) = Pr(\Psi \rightarrow \Phi)$ then, since $\Psi \rightarrow \Phi$ is replaceable in the latter statement by its contrapositive, we could accommodate the present case. But **ProbCal** does not license the equation of $Pr(\Phi/\Psi)$ and $Pr(\Psi \rightarrow \Phi)$.

In all three cases, we seem to have in play a legitimate notion of probability, one moreover that captures an idea of a proposition's degree of certainty depending on the available evidence. It looks, therefore, as if one *can* speak of inductive probability. But, note well, if inductive probability is genuinely at issue here, the coincidence between inductive probability and aleatory probability is broken; that is, there are some cases of genuine inductive probability which don't obey the **ProbCal** axioms. Following Cohen, let us call this *Baconian probability.* And we see at once, the necessity of now regarding inductive probability as a genus of which Baconian probability and aleatory or Pascalian probability are at best species.

The existence and apparent legitimacy of Baconian probability presents the inductive logician with a complication. If probability is retained as a significant conceptual constituent of the very idea of induction, then the theorist is faced with the task of finding a principled method for determining whether a real-life evidentiary situation is one which models inductive probability of the Baconian or of the Pascalian ilk. The theoretical recalcitrance of such a task is perhaps nowhere better evidenced than in contemporary theoretical jurisprudence in which it is hotly debated whether legal evidence, when held to a balance of probability standard, is susceptible to deep Bayesian description. (Tillers and Green [1988]).

The introduction of Baconian probability into the logician's maw also makes possible an interesting and important philosophical contention. There is an abundance of experimental data which strongly suggests that human reasoners have a hefty disposition to commit certain kinds of statistical and probabilistic errors. Many psychologists are of the view that such errors satisfy the definition of *fallacy*; that is, that they are errors which are:

(1) *attractive* and easy to make;

(2) *universal* or made cross-culturally by lots of people

(3) *incorrigible* or hard to correct.

Some psychologists go even further in their interpretation of these results. Why, they ask, would an error satisfy condition (2)? Surely because human beings are *disposed to commit* such errors. If so, conditions (1) and (3) become hardly more than minor codicils to the main dispositional claim.

As against this, Cohen [1979c], [1980c], [1981], and [1994], it has been argued that, although the experimental results clearly indicate what appears to be pervasive violations of **ProbCal**, these violations, count as errors only if the probabilistic problems which the experimenters sets for these subjects required the application of Pascalian probability. But, as the Vermeer example clearly indicates, sometimes there is probabilistic reasoning which is Pascalianly bad and Baconianly good in contexts, furthermore, in which Baconian probability, not Pascalian, is the appropriately applicable probability concept. In such cases, conditions (1), (2) and (3) might well be satisfied. But since there is no concomitant *error*, these would not, to say the least, be fallacies.

This leaves us with some interesting questions: Is there a theory of Baconian probability that even begins to promise the level of theoretical sophistication of **ProbCal** for Pascalian probability? Granted the distinction between the two, is there a principled way of determining the appositeness of one over the other in arbitrarily selected task-settings for human agents. Given the conceptual appositiveness of Pascalian probability to certain types of reasoning tasks, how realistic is it to suppose that beings like us can actually conform their reasonings to the strictures of **ProbCal**?

We leave the first two of these questions for consideration elsewhere. The third we shall deal with briefly, having met with it previously. Recall that a hallmark of Bayesianism is that (Pascalian) reasoning is (or ought to be) conditionalization. That is, in those cases in which it is appropriate for a reasoner to apply the Pascalian concept of probability, then it is correct for him to revise his belief sets in the following manner. Let Φ be a statement and $\sigma_1, \ldots, \sigma_n$ the simple evidence statementsrelevant to Φ. As before, where σ_i is an atomic statement, $\pm\sigma_i$ is a literal, and a χ_i is a complete inventory, i.e., a complete conjunction of literals. Since there are n atomic evidence-statements, there are 2^n complete inventories. Then the new probability that a reasoner does or should assign to Φ is the sum of all the following numbers for each i:

$$newPr(\chi_i)x\frac{oldPr(\Phi \wedge \Phi_i)}{oldPr(\chi_i)}$$

The number of complete inventories increases exponentially as the number of potentially relevant evidence-statements. This is discouraging. For, to

> accept or reject any ten evidence propositions, one would have to record probabilities of over a thousand such conjunctions for each proposition one is interested in updating .... For twenty evidence propositions, one must record a million probabilities. (Harman [1986, 26]. See also chapter 4 below.)

The fact of such exponential explosions puts enormous pressure on the idea that in those situations in which Pascalian probability is appropriately applied, the reasoner's belief-revisions will be sound to the extent that one conforms his revisions to **ProbCal**. Human beings can't conform, except in extremely simple cases, since they can't manage the computational explosion of even modestly realistic and simple cases.

The factor of exponential explosion is occasion to revisit deductive logic. As we have seen, some logicians are of the view that deductive logic is a good theory of inference or belief-revision. If this is right, then a deductive reasoner should conform his revisions to the requirements of systems such as **PC** and **PredCal**. In such systems, everything follows from an inconsistency; otherwise not everything follows. One might think that, at a minimum, the deductive reasoner has a stake in knowing, at given times, whether his belief set is truth functionally consistent. Any reasoner for whom this is a recognized question has put to himself the *satisfiability problem*. The satisfiability problem is an *NP*-problem, a nondeterministic polynomial time problem, i.e., a problem solvable on a nondeterministic Turing machine in polynomial time. Any *NP*-problem is known to reduce to the satisfiability problem, which thus may stand in for them all: or, as complexity theorists like to put it, the satisfiability problem is NP*-complete*. (See section 2.11 of the previous chapter.)

We have also seen that P is the class of all computational problem classes computable in deterministic polynomial time. Although it has not been decisively proved, it is accepted that $NP \neq P$.

It is easy to see that a P-problem is computationally tractable, that is, can be solved without an exponential explosion. This matters for *NP*-problems. Are they computationally tractable, in the sense that the procedures that solve them have an execution time which increases as a polynomial function of the number of inputs it processes? The answer is that in the relevant sense they are tractable problems. However, *NP*-problems differ from P-problems only when the "solver" has limitless capacity for (certain) *parallel* computations. No human reasoner possesses anything like this capacity. So in a practical sense, *NP*-problems are intractable, and they beset their would-be solvers with the unanswerable vexations of something similar to exponential explosion. Thus no human reasoner can check his belief-set for consistency, except for belief-sets of unrealistically small memberships.

We seem to have it that no human being could be a model of a type of rationality based on deductive logic.

ACKNOWLEDGEMENT

As with the prior chapter, Professor Erik C. W. Krabbe has supported the writing of this chapter with his remarkable critical generosity.

BIBLIOGRAPHY

[Atkinson and Peijnenburg, 1999] David Atkinson and Jeanne Peijnenburg. Probability as a theory dependent concept, *Synthese*, 118:307–328, 1999.

[Arnauld and Nicole, 1996] Antoine Arnauld and Pierre Nicole. *Logic or the Art of Thinking*. Cambridge: Cambridge University Press, 1996. Originally published in 1662. This volume edited by Jill Vance Buroker.

[Bacon, 1905] Francis Bacon. Novum organum. In R.L. Ellis and J. Spedding, editors, *The Philosophical Works of Francis Bacon*, pages 212–387. London: Routledge, 1905.

[Billingsley, 1979] Patrick Billingsley. *Probability and Measure*. New York: John Wiley, 3rd edition, 1979.

[Burks, 1977] Arthur W. Burks. *Cause, Chance, Reason*. Chicago: University of Chicago Press, 1977.

[Carnap, 1949] Rudolf Carnap. The two concepts of probability. In Herbert Feigl and Wilfrid Sellars, editors, *Readings in Philosophical Analysis*, pages 330–348. New York: Appleton-Century-Crofts, 1949.

[Carnap, 1952] Rudolf Carnap. *The Continuum of Inductive Methods*. Chicago; University of Chicago, 1952.

[Carnap, 1962] R. Carnap. *Logical Foundations of Probability*. Chicago: University of Chicago Press, 2nd edition, 1962.

[Cohen, 1979c] L. Jonathan Cohen. On the psychology of prediction: Whose is the fallacy? *Cognition*, 7:385–407, 1979c.

[Cohen, 1980c] L. Jonathan Cohen. Some historical remarks on the Baconian conception of probability. *Journal of the History of Ideas*, 41:219–231, 1980c.

[Cohen, 1981] L. Jonathan Cohen. Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences*, 4:317–331, 1981.

[Cohen, 1989] L. Jonathan Cohen. *An Introduction to the Philosophy of Induction and Probability*. Oxford: Clarendon Press, 1989.

[Cohen, 1994] L. Jonathan Cohen. Some steps towards a general theory of relevance. *Synthese*, 101:171–185, 1994.

[de Finetti, 1964] Bruno de Finetti. Foresight. Its logical laws, its subjective sources. In Henry Kyburg and Howard Smakler, editors, *Studies in Subjective Probability*. New York: John Wiley, 1964.

[Earman, 1992] John Earman. *Bayes or Bust*. Cambridge, MA: MIT Press, 1992.

[Gabbay and Woods, 2002] Dov M. Gabbay and John Woods. *The Reach of Abduction*. To appear.

[Goodman, 1983] Nelson Goodman. *Fact, Fiction and Forecast*. Cambridge: Harvard University Press, 4th edition, 1983.

[Gustason, 1994] William Gustason. *Reasoning From Evidence*. New York: MacMillan, 1994.

[Hacking, 1975] Ian Hacking. *The Emergence of Probability*. Cambridge: Cambridge University Press, 1975.

[Harman, 1986] Gilbert Harman. *Change in View*. Cambridge, MA: MIT Press, 1986.

[Hempel, 1945] C.G. Hempel. Studies in the logic of confirmation. *Mind*, 54:1–26 and 97–121, 1945. Reprinted in C.G. Hempel *Aspects of Scientific Explanation* by the Free Press of New York in 1965.

[Hintikka, 1966] Jaakko Hintikka. A two-dimensional continuum of inductive methods. In *Aspects of Inductive Logic*, pages 113–132. Amsterdam: North-Holland, 1966.

[Hosiasson-Lindenbaum, 1940] J. Hosiasson-Lindenbaum. On confirmation. *Journal of Symbolic Logic*, 5:133–148, 1940.

[Howson, 1995] Colin Howson. Theores of probability. *British Journal for Philosophy of Science*, 46:1–32, 1995.

[Hume, 1902] David Hume. *Enquiries concerning the Human Understanding and concerning the Principles of Morals.* Oxford: Clarendon Press, 1902. Reprinted from the posthumous edition of 1777 and edited by L.A. Selby-Bigge. 2nd Edition.

[Jeffrey, 1983] R. Jeffrey. *The Logic of Decision.* Chicago: University of Chicago Press, 2nd edition, 1983. First published in 1965.

[Jeffreys, 1939] Harold Jeffreys. *Theory of Probability.* Oxford: Oxford University Press, 1939.

[Johnson, 1932] W.E. Johnson. Probabilty. *Mind*, 41:409, 1932.

[Kemeny, 1963] J.G. Kemeny. Analyticity versus fuzziness. *Synthese*, 15:57–80, 1963.

[Keynes, 1921] John Maynard Keynes. *A Treatise in Probability.* New York and London: Harper and Row and Macmillan, 1921.

[Kolmogorov, 1950] A.N. Kolmogorov. *The Fundations of the Theory of Probability.* New York: Chelsea, 1950.

[Kolmogorov, 1956] A. Kolmogorov. *Foundations of the Theory of Probability.* London: Chelsea, 1956. Originally published in 1933. This is the second English edition.

[Lewis, 1980] David K. Lewis. A subjectivist's guide to objective chance. In David K. Lewis, editor, *Philosophical Papers*, volume 2, pages 83–132. Oxford: Oxford University Press, 1980.

[Lipton, 1991] Peter Lipton. *Inference to the Best Explanation.* London: Routledge, 1991.

[Mellor, 1971] David H. Mellor. *The Matter of Chance.* Cambridge: Cambridge University Press, 1971.

[Popper, 1962] K. Popper. *Conjectures and Refutations.* New York: Basic Books, 1962.

[Popper, 1992] Karl Popper. *The Logic of Scientific Discovery.* New York: Routledge, 1992. Originally published in 1934.

[Putnam, 1975] Hilary Putnam. Degree of confirmation and inductive logic. In Hilary Putnam, editor, *Mathematics, Matter and Method*, pages 270–292. Cambridge: Cambridge University Press, 1975.

[Putnam, 1975b] Hilary Putnam. Probability and confirmation. In Hilary Putnam, editor, *Mathematics, Matter and Method*, pages 293–304. Cambridge: Cambridge University Press, 1975b. Originally published in *The Voice of America, Forum Philosophy of Science* volume 10 in 1963.

[Ramsey, 1931] Frank Ramsey. Truth and probability. In Frank Ramsey, editor, *The Foundations of Mathematics and Other Essays*, pages 156–198. London: Routledge and Kegan Paul, 1931. originally published in 1926.

[Rawls, 1971] John Rawls. *A Theory of Justice.* Cambridge MA: Harvard University Press, 1971.

[Rogers, 1967] Hartley Rogers. *Theory of Recursive Functions and Effective Computability.* New York: McGraw-Hill, 1967.

[Rosenkrantz, 1981] R.D. Rosenkrantz. *Foundations and Applications of Inductive Probability.* Atascadero, CA: Ridgeview, 1981.

[Shimony, 1955] Abner Shimony. Coherence and the axioms and confirmation. *Journal of Symbolic Logic*, 20:1–28, 1955.

[Stalker, 1994] Douglas Stalker. *Grue!: The New Riddle of Induction.* Chicago: Open Court, 1994.

[Tillers and Green, 1988] Peter Tillers and Eric D. Green. *Probability and Inference in the Law of Evidence.* Dordrecht and Boston: Kluwer, 1988.

[von Mises, 1928] Richard von Mises, editor. *Probability, Statistics and Truth.* London: Fuller and Unwin, 1928.

[von Mises, 1964] Richard von Mises. *Mathematical Theory of Probability and Statistics.* New York: Academic Press, 1964.

[Watkins, 1957] J.W.N. Watkins. Between analytic and empirical. *Philosophy*, 32:112–131, 1957.

[Woods, 2001e] John Woods. *Death of Argument: Fallacies and Other Seductions*. Newport Beach: Vale Press, 2001e. In press.

GILBERT HARMAN

INTERNAL CRITIQUE: A LOGIC IS NOT A THEORY OF REASONING AND A THEORY OF REASONING IS NOT A LOGIC

In order to understand the relations between reasoning and logic, it is crucial not to confuse issues of implication with issues of inference. Inference and implication are very different things and the relation between them is rather obscure. Implication is a fairly abstract matter, a relation among propositions. Inference and reasoning are psychological processes, processes of reasoned change in view (or of reasoned no change in view).

Logic as a theory of implication is a very different sort of theory from logic as a theory of reasoning or methodology. Historically the term "logic" has been used in both ways. Current usage favours restriction of the term "logic" to the theory of implication. The theory of reasoning is best called "the theory of reasoning" or "methodology".

If there are any principles of inference or reasoning, they are normative principles about when it is rational or reasonable to reach a certain conclusion. Principles of implication are not normative (outside of deontic logic) and do not have a psychological subject matter (outside of the logic of belief).

Implication is relatively well understood. There are many technical studies of implication and of logic understood as the theory of implication. Inference and reasoning are not well understood. This is because a theory of inference or reasoning must be part of a theory of rationality and rationality is not well understood.

The present chapter discusses relations between the study of reasoning and the theory of implication.

1 THEORETICAL AND PRACTICAL RATIONALITY

It is traditional to distinguish theoretical reasoning, which most directly affects beliefs, from practical reasoning, which most affects plans and intentions. Theoretical reasoning is reasoned change in belief, or reasoned no-change in belief. Practical reasoning is reasoned change in plans and intentions, or reasoned no-change.

There are many similarities between theoretical and practical reasoning, but there are also important differences. One difference is that a certain sort of arbitrary choice is permitted in practical but not in theoretical reasoning. Given a choice among several equally acceptable things to do, it can be rational arbitrarily to choose one and irrational not to make this arbitrary choice. But, given a choice among several equally acceptable beliefs, it is

never rational arbitrarily to select one to believe. Rationality requires a suspension of judgement in that case.

Someone who is unable to make arbitrary choices of things to do suffers from a serious defect in practical rationality, like Buridan's ass, stuck halfway between equally attractive equally delicious piles of hay. With belief it is just the opposite.

Consider also wishful thinking. It is theoretically unreasonable, but practically reasonable. One's goals, wishes, and desires are relevant to practical reasoning in a way in which they are not relevant to theoretical reasoning. A desire for more money can rationally influence one's decision to take a better paying job. Such a desire should not rationally influence one's conclusion, reached after having been interviewed for the job, that the interview went well and one will be offered the job. It is irrational to let one's desire to have done well at the interview influence one's conclusion about whether or not one did well. To believe that something is so merely because one wants it to be so is theoretically unreasonable, whereas to decide to try to make something so because one wants it to be so is reasonable practical thinking. Desires can rationally influence the conclusions of practical reasoning in a way that they cannot rationally influence the conclusions of theoretical thinking.

On the other hand, one's goals are in other ways relevant to one's theoretical reasoning. In particular, one's goals may give provide a very strong reason to think about one thing rather than another. One's desires help to set the problems that one's reasoning is aimed at resolving. It is overly simple to say that one's desires cannot rationally affect what conclusions are legitimately reached in theoretical reasoning. One's desires can rationally affect one's theoretical conclusions by affecting what questions one uses theoretical reasoning to answer.

The point about wishful thinking then is this: *given what question one is using theoretical reasoning to answer*, one's desires cannot rationally affect what answer one reaches to that question, whereas in practical reasoning one's desires can rationally influence not just the questions one considers but also what practical answers one gives to those questions.

1.1 Legitimate practical reasons for beliefs

There are complications. Sometimes one has good practical reasons to answer certain questions in a certain way. There might be evidence that people who believe they will recover quickly from a certain illness are more likely to recover quickly than are people who are otherwise the same but do not believe they will recover quickly. A person with that illness could then have a practical reason to believe that he or she will recover quickly. Similarly, a contestant in a tennis tournament may have a practical reason to believe that he or she will win the tournament, if there is evidence that posses-

sion of such a belief improves one's playing. Or, to take a slightly different example, a salesperson might make more sales if he or she believed in the value of the item to be sold; so such a salesperson might have a practical reason to have such a belief.

There are other cases as well. One may have a practical reason of loyalty to believe that one's friend is not guilty of a crime with which he or she has been charged. One may want to have certain beliefs in order to fit in with the "in crowd" or to obtain a job that will go only to someone with the relevant beliefs. In these and various other cases there may be good practical reasons to believe something.

This indicates that the difference between practical reasons and theoretical reasons is not just a matter of what they are reasons for—intentions versus beliefs. The difference has to do with the way in which reasons are reasons.

Some of the examples just discussed mention a reason to believe something that does not make it more likely that the belief is true. Such reasons are sometimes called (e.g., by Foley, 1987) non-epistemic reasons for belief, in contrast with the more usual epistemic reasons for belief that do make a belief more likely to be true. We will see below that not all practical reasons for belief are non-epistemic. Some practical reasons do make a belief more likely to be true.

1.2 Inference versus implication (again)

I now want to return to the point with which this chapter began, namely that issues about inference and reasoning need to be distinguished from issues about implication and consistency. Inference and reasoning are psychological processes leading to possible changes in belief (theoretical reasoning) or possible changes in plans and intentions (practical reasoning). Implication is more directly a relation among propositions. Certain propositions imply another proposition when and only when, if the former propositions are true, so is the latter proposition.

It is one thing to say "A, B, and C imply D". It is quite another thing to say, "If you believe A, B, and C, you should or may infer D". The first of these remarks is a remark about implication. The second is a remark about inference. The first says nothing special about belief or any other psychological state, unless one of A, B, or C has psychological content, nor does the first remark say anything normative about what anyone "should" or "may" do [Goldman, 1986].

The first remark can be true without the second being true. It may require considerable logical talent, even genius, to see that the implication holds. A person without such genius or talent may believe A, B, and C without having any reason at all to believe D.

Furthermore, a person who believes A, B, and C and realises that A, B, and C imply D may also believe for very good reason that D is false. Such a person may now have a reason to stop believing one of A, B, or C rather than a reason to believe D.

Moreover, even someone who believes A, B, and C, who realises that A, B, and C imply D, and who has no reason to think that D is false, may have no reason to infer D. Such a person may be completely uninterested in whether D is true or false and no reason to be interested.

Many trivial things follow from one's beliefs without one having any reason to infer them. One has no reasons to clutter one's mind with trivialities just because they follow from other things one believes.

These and related examples indicate that the connection between inference and implication is fairly complex. We will say more about it below.

Similar remarks hold for consistency. Just as issues about implication have to be distinguished from issues about reasonable inference, issues about consistency have to be distinguished from issues about rationality and irrationality. Consistency and inconsistency are in the first instance relations among propositions and only indirectly relations among propositional attitudes. Propositions are consistent when and only when it is possible for them all to be true together. Propositions are inconsistent when and only when it is not possible for them all to be true together.

So, it is one thing to say that certain propositions are inconsistent with each other and quite another to say that it is irrational for someone to believe those propositions. The first remark, unlike the second, says nothing special about belief or other psychological states, nor does it say anything normative. So, the first remark can be true without the second being true. Suppose one believes each of the propositions. The inconsistency may have gone unnoticed and may be very difficult to discover. One's failure to discover the contradiction need not indicate any irrationality.

And, even if one notices that the beliefs in question are inconsistent, one may still have reasons to continue to accept each and it may be quite unclear which should be given up. One may not have the time or the ability to work out which should be given up. One may have more urgent matters to attend to before resolving this inconsistency in one's beliefs. One may be very hungry and want to have lunch before solving this problem. In the meantime, it may very well be rational for one to continue to believe them all.

1.3 Ideal reasoners?

Reasoning is subject to resource limits of attention, memory, and time. So, it is not rational to fill your time inferring trivial consequences of your beliefs when you have more important things to attend to. And you cannot

be charged with irrationality for having inconsistent beliefs where it would be costly to avoid the inconsistency.

Some theories of rationality [Stalnaker, 1984] abstract away from resource limits. Such theories of idea rationality are concerned with an "ideally rational agent" whose beliefs are always consistent and closed under logical implication. Other theorists argue that such an idealisation appears to confuse rationality, ideal or otherwise, with logical genius and even divinity! Furthermore, as we shall see, it is unclear how to relate such an "ideal" to actual finite human beings with their resource-limited rationality.

We have already seen that ordinary rationality requires neither deductive closure nor consistency. Ordinary rationality does not require deductive closure, because one is not always rational to believe something simply because it is implied by one's other beliefs. Rationality does not require consistency, because one can be rational even though there are undetected inconsistencies in one's beliefs, and because it is not always rational to respond to the discovery of inconsistency by stopping whatever else one might be doing in order to eliminate the inconsistency.

Now consider an ideal agent with no limitations on memory, attention span, or time, with instantaneous and cost-free computational abilities. It is not obvious whether such an agent would have a reason to infer all the trivial consequences of his or her beliefs. Although it would not cost anything for the agent to draw all those consequences, even all infinitely many of them, there would also be no need to draw any of those consequences in the absence of a reason to be interested in them, since the agent can effortlessly compute any consequence whenever it might be needed.

Could an ideal agent's beliefs be inconsistent? Suppose ordinary classical logic. Then, if the agent's beliefs were also deductively closed, the agent would then believe everything, because everything follows from inconsistency in classical logic.

This raises a question as to how the ideal agent could recover from inconsistency. Ordinary rational agents deal with momentary inconsistency all the time. One believes P but discovers Q, realising that P and Q cannot both be true. One believes that something will happen and is surprised when it does not happen. For a moment, one has inconsistent beliefs. Normally one quickly recovers from the inconsistency by abandoning one of the initial beliefs.

But consider the implications of this sort of surprise for an ideal deductively closed agent. If the beliefs of such an agent were even momentarily inconsistent, the agent could never rationally recover, since there would be no trace in the agent's beliefs of how the agent had acquired the inconsistent beliefs. Since rational recovery from inconsistency can appeal only to present beliefs, and, since the deductively closed agent has exactly the same beliefs no matter how he or she got into inconsistency, there is no way in which the deductively closed agent could use temporal criteria in retreating

from inconsistency—the agent would have to recover in exactly the same way, no matter where he or she had started.

It is therefore quite unclear how ideal rational agents might deal with ordinary surprise. Various possibilities suggest themselves, but we need not consider them, because in what follows we will be directly concerned with real rather than ideal rational agents.

2 GENERAL CONSERVATISM VERSUS SPECIAL FOUNDATIONS

Ordinary reasoning is conservative in the sense that one starts where one is, with one's current beliefs and intentions. Rational and reasonable change in view consists in trying to make improvements in one's initial position. One's initial beliefs and intentions have a privileged position in the sense that one begins with them rather than with nothing at all or with some special privileged part of those beliefs and intentions. So, for example, one ordinarily continues to believe something that one starts out believing in the absence of a special reason to doubt it.

An alternative and radical conception of rationality going back to Descartes (1637) requires beliefs to be associated with reasons or justifications. Such justifications appeal to other beliefs, themselves to be associated with justifications, and so forth, until certain special foundational beliefs are reached that are self-justifying and need no further justification. Special foundational beliefs include beliefs about immediate experiences, such as headaches and perceptions, obvious logical and mathematical axioms, and other intuitively obvious truths. Rational change in view and rational no-change in view are to start always from evidence—those propositions that are evident. Then one is to accept only what can be justified from one's evidence basis, in this view.

Recent versions of special foundationalism [Foley, 1987; Alston, 1989; Chisholm, 1982] do not require foundational beliefs to be guaranteed to be true. In the absence of special challenges to them, they are justified, but their initial justified status might be overridden by special reasons to doubt them.

Supposing that there are two conflicting theories of reasoning, special foundationalism and general conservatism, each can be described using the terminology of the other theory as follows: the special foundations theory is conservative about all foundational beliefs, but only foundational beliefs, whereas general conservatism treats all beliefs as foundational.

One problem for special foundationalism is to explain why the special foundational beliefs should have the sort of special status assigned to them in special foundationalism. What distinguishes foundational beliefs from other beliefs in such a way that justifies conservatism with respect to the foundational beliefs but not the other beliefs?

A second and potentially more serious problem is that people simply do not keep track of their reasons for their non-foundational beliefs. This is a problem because, according to special foundationalism, if one does not associate a complete enough justification with a given non-foundational belief, then it is not rational or reasonable for one to continue to believe it. This may undermine a great many of one's beliefs. Few people can remember their reasons for various of their beliefs about geography or history; does that make it unreasonable for them to continue to believe, for example, that Rome is in Italy or that the Battle of Hastings was in 1066? Furthermore, when beliefs are acquired on the basis of perception, one rarely continues to remember the perceptual evidence on which the beliefs were based. If that makes it irrational to believe, for example, that one saw John yesterday, the reasonable thing to do is to abandon almost everything one believes!

The issue between the two approaches to reasoning amounts to a question about the burden of proof or justification. According to special foundationalism, the burden of justification falls on continuing to believe something, at least for non-foundational beliefs. Any non-foundational belief requires special justification.

Foundational beliefs do not require special justification. For them, what requires justification is failing to continue to believe them. Sometimes there is a reason to abandon a foundational belief, but such abandonment requires such a special reason.

According to general conservatism, the burden of justification is always on changing beliefs or intentions. One starts with certain beliefs and intentions and any change in them requires some special reason. any sort of change in belief or intention requires special justification. Merely continuing to believe what one believes or intends requires no special justification in the absence of a special challenge to that belief or intention.

Clearly general conservatism fits better with ordinary thinking. Special foundationalism would imply that it is irrational or unreasonable for a typical person to continue to believe most of what he or she believes.

3 INDUCTION AND DEDUCTION

Some authors draw a mistaken contrast between deductive and inductive reasoning. This is a mistake. Deduction and induction are not two kinds of reasoning; they are not two kinds of anything.

Deduction is concerned with certain relations among propositions, especially relations of implication and consistency. Induction is not concerned with those or any similar sort of relation. Induction is a kind of reasoning; but deduction is not a kind of reasoning.

Deductive logic is sometimes presented via a certain notion of "proof" or "argument". A proof or argument in this sense has premises, intermediate

steps, and a final conclusion. Each step must follow logically from prior steps in accordance with one or another specific rule, sometimes (misleadingly) called a "rule of inference". Sometimes a proof or argument is taken to be an instance of "deductive reasoning." Deductive reasoning in this sense is sometimes contrasted with "inductive reasoning", which allegedly takes a similar form, with premises, intermediate steps, and final conclusions, but with the following difference: deductive steps are always truth preserving, whereas inductive steps are not.

This way of looking at deduction and induction is very misleading. For one thing, consider the reasoning that goes into the construction of a deductive proof or argument. Except in the simplest cases, the best strategy is not to expect to start with the premises, figure out the first intermediate step of the proof, then the second, and so on until the conclusion is reached. Often it is useful to start from the proposition to be proved and work backward. It is also useful to consider what intermediate results might be useful.

In other words, the so-called deductive rules of inference are not rules that you follow in constructing the proof. They are rules that the proof must satisfy in order to be a proof!

There is a difference between reasoning about a proof, involving the construction of a proof that must satisfy certain rules, and reasoning that proceeds temporally in the same pattern as the proof in accordance with those rules. One does not reason deductively in the sense that one reasons in the pattern of a proof. One can reason about a deductive proof, just as one can reason about anything else. But one's reasoning is not well represented by anything like a proof or argument in the above sense.

3.1 Deduction

Deduction is not a kind of inference or reasoning, although one can reason about deductions. Deduction is implication. A deduction or proof or argument exhibits an implication by showing intermediate steps.

Logic, conceived as the theory of deduction, is not by itself a theory of reasoning. In other words, it is not by itself a theory about what to believe or intend. It is not a theory concerning reasoned change in view or reasoned no change in view.

It is true that deductions, proofs, and arguments do seem relevant to reasoning. It is not just that you sometimes reason about deductions in the way you reason about your finances or where to go on your summer trip. It is an interesting and nontrivial problem to say just how deductions are relevant to reasoning, a problem that is hidden from view by talk of deductive and inductive reasoning, as if it were obvious that some reasoning follows deductive principles.

It may be useful to construct deductions in some reasoning about ordinary matters, and not just when one is explicitly reasoning about deductions or

proofs. But why should it be useful to construct deductions? What role do they play in reasoning?

Sometimes one accepts a conclusion because one has constructed a proof of that conclusion from other things one accepts. But there are other cases in which one constructs a proof of something one already accepts in order to see what assumptions might account for it. In such a case, the conclusion that one accepts might be a premise of the proof. The connection between proofs and reasoning is therefore complex.

3.2 Induction

The term "induction" is sometimes is sometimes restricted to "enumerative induction" in which a generalisation is inferred from its instances. But the term "induction" is often used more widely so as to include inference to the best explanation of one's evidence. What makes one hypothesis "better" than another for the purpose of inference to the best explanation is an issue that we must discuss below.

Philosophers sometimes discuss a "problem of induction" [BonJour, 1992]: How can one be justified in drawing a conclusion that is not guaranteed to be true by one's premises? But it is unclear what the problem of induction is supposed to be. Premises of an argument are to be distinguished from the starting points in reasoning, as I have already observed. The conclusion of an argument is not to be identified with the conclusion of reasoning, in the sense of what you end up with or "conclude" as the result of your reasoning. Even when reasoning culminates in the construction of an argument, the conclusion of the *argument* may be something one started off believing, and the conclusion of one's *reasoning* may be to accept something that is a premise of an explanatory argument constructed as the result of inference to the best explanation.

Clearly, it would be stupid, indeed highly irrational, not to engage in inductive reasoning. One would no longer be able to learn from experience. One would have no basis for any expectations at all about the future, since one's evidence entirely concerns the past.

The "problem of induction" is a creation of confusion about induction and deduction, arising from the deductive model of inference. Again, it is important to see that there are not two kinds of reasoning, deductive and inductive. Deduction has to do with implication and consistency and is only indirectly relevant to what one should believe.

4 COHERENCE

Everything one believes is at least potentially relevant to the conclusions one can reasonably draw. Rationality is a matter of one's overall view, including one's beliefs and one's intentions.

If it is reasonable to change one's view in a certain way, we might say that one's view would be more rationally "coherent" if changed in that way. We can describe principles of rationality as principles of rational coherence. Adopting this terminology, we can (following Pollock, 1974) distinguish two sorts of coherence, positive and negative.

Negative coherence is merely the absence of incoherence. Beliefs and intentions are incoherent to the extent that they are inconsistent with each other or clash in other ways. Incoherence is something to be avoided, if possible, although I have observed that it is not always possible to avoid incoherence. One's beliefs might be inconsistent without one's realizing that they are. And, even if one is aware of inconsistency, one may not know of a sufficiently easy way to get rid of it. Still, to the extent that one is aware of incoherence in one's view, one has a reason to modify one's view in order to get rid of the incoherence, if one can do so without too much expense.

Here, then, is one way that logic, in the sense of the theory of deductive implication, might be relevant to the theory of rationality, through providing an account of (one kind of) incoherence or inconsistency.

Positive coherence among one's beliefs and intentions exists to the extent that they are connected in ways that allow them to support each other. Relevant connections may involve explanations, generalisations, and implications.

In many cases new conclusions are accepted because they are part of explanations that serve to connect and integrate prior beliefs. A detective tries to find the best explanation of the various clues and other evidence. Scientific theories are accepted because of the way they allow us to explain the phenomena. Acceptance of someone's testimony may involve acceptance of certain explanations, namely that the speaker's testimony is the result of a desire to say what's true plus certain beliefs about what is true, where these beliefs are the result of the speaker's being in a position to know what is true. When a car fails to start, one may infer that the battery is low: that best explains the disappointing sounds that occur when the key is turned.

In some cases generalisations are accepted on the basis of a prior acceptance of certain instances. We might treat this as a special case of inference to the best explanation, supposing that the accepted generalisations explain their instances. But then we must recognise that this is a different sort of explanation from the causal or quasi-causal explanation involved in accounting for testimony or the failure of a car to start. A general correlation does not cause its instances.

Implication is also an important connector. Sometimes one does accept

a new conclusion because it is implied by things one already accepts. This is a second way in which deductive logic, as a theory of deductive relations, can be relevant to reasoning. Deductive logic, so construed, is a theory of deductive implication and implication can be a coherence giving connection.

In trying to develop a theory of rational coherence (something that does not yet exist), we might try to reduce some of these coherence giving factors to others. For example, we might try to reduce all cases to explanatory coherence. But that is implausible for many cases in which a conclusion is accepted because it is implied by other beliefs. What is the relevant explanation? We might say that the argument leading to the conclusion explains "why the conclusion is true". But that seems to stretch the notion of explanation.

Another idea would be to try to reduce all coherence to that involved in implication. That has some plausibility for certain explanations. And strict generalisations are related to their instances by implication. Explanations in physics often seem to work via implication, leading to the "deductive nomological model" of explanation [Hempel, 1965]. But not all explanations take this form. Many explanations appeal to "default" principles that hold only "other things being equal" or "normally". Such explanations are not easily treated as deductively valid arguments.

5 SIMPLICITY

In trying to explain some data, it is reasonable to consider a very limited range among the infinitely many logically possible explanations. The reasonable inquirer restricts attention to the set of relatively simple hypotheses that might account for most of the data.

This is not to say very much, since it amounts to using the term "simple" for whatever the relevant factors are that restrict rational attention to a certain few hypotheses. Furthermore, we are concerned with *relative* simplicity in this sense. A hypothesis that is too complicated as compared with other available hypotheses at one time can have a different status at another time if those other hypotheses have been eliminated. The first hypothesis might then be among the simplest of available hypotheses.

So, to say that the rational inquirer is concerned to find a simple hypothesis is not to say that the rational inquirer is committed to believing that "reality is simple," whatever that might mean.

Goodman [1965] discusses a classic example. Given that all emeralds examined up until now have been found to be green, the evidence supports the hypothesis that all emeralds are green. One would not normally even consider the competing hypothesis that all emeralds are either green if first examined before A.D. 2100 or blue if not first examined before A.D. 2100. Given a suitable definition of "grue," the latter hypothesis can be written,

"All emeralds are grue". The grue hypothesis conflicts with the first hypothesis as regards any emeralds not first examined by A.D. 2100. According to the first hypothesis those emeralds are green; according to the second they are blue.

Goodman points out the hypotheses like the second, grue hypothesis are not taken seriously. His "new riddle of induction" asks what the difference is between hypotheses like the first hypothesis that are taken seriously and hypotheses like the grue hypothesis that are not taken seriously. Clearly, there is a sense in which the answer has something to do with simplicity. The first hypothesis is much simpler than the second. But what sort of simplicity is in question and why should it be relevant?

In thinking about this it is very important to see that using simplicity to rule hypotheses out of consideration is to be distinguished from using simplicity as an explicit consideration in theory choice. Sometimes a scientist will say that a particular theory is better than another because the first theory assumes the existence of fewer objects, fewer basic principles, or whatever. When a scientist argues in some such way, he or she is arguing in favour of one rather than another hypothesis that is already being taken seriously. As Sober [1988] has observed, such appeals to simplicity are often quite controversial. That is, it is controversial whether simplicity in one or another respect is a relevant consideration in choosing among hypotheses.

But even where there are deep controversies in a subject, reasonable disputants will still take seriously only a very few of the infinitely many possible hypotheses. We are concerned with whatever it is that leads reasonable people to disregard most of the hypotheses as too "silly" to be considered (remembering that silliness is a relative matter: we can imagine circumstances in which previously ignored hypotheses become discussable).

Let us call the sort of simplicity we are concerned with "basic simplicity". Since crazy or silly hypotheses are ruled out in all domains, let us assume that there is a single domain independent notion of basic simplicity.

The basic simplicity of a hypothesis seems to have something to do with the simplicity of its representation. But it is always possible to represent any hypothesis simply, so the matter is a bit more complex. Using the new term "grue" Goodman's second hypothesis can be represented as simply as the original hypothesis that all emeralds are green. In fact, any hypothesis can be abbreviated by a single symbol, so simplicity of surface representation cannot be taken at face value.

But suppose a hypothesis like, "All emeralds are green," is used to explain the data. Then it has to be expanded to its more complex form, "All emeralds are either green if first examined before A.D. 2000 or blue if not first examined before A.D. 2000". This expansion is needed on the assumption that we are more interested in accounting for the colors of objects, such as whether they are blue or green, as opposed to their "cholers", such as whether they are grue or bleen. If instead we were more interested

in explaining why emeralds were grue, we could use the hypothesis, "All emeralds are grue", without having to expand it, and the hypothesis that "All emeralds are green", would require elaboration in terms of grue and bleen in order to provide the desired explanation.

So, perhaps the thing to look at is not so much the mere statement of the hypothesis but also how complicated it is to use the hypothesis to explain the data and predict new observations of a sort in which we are interested. In considering possible explanations of given data, it appears to be rational and reasonable to ignore hypotheses that are much harder to use in explanation and prediction than other available hypotheses that in other respects account equally well for the data. This represents a further respect in which theoretical rationality depends on practical concerns.

6 RATIONALITY IN ACTION

We have seen that practical considerations play a number of roles in theoretical rationality. Practical considerations can help determine what questions it is rational to try to answer. Furthermore, practical considerations may play a role in determining what hypotheses theoretical rationality takes seriously and may also be relevant to the conservatism of theoretical reasoning. But I have not said much about practical rationality except to observe that it is often rational to make an arbitrary choice about what to do in a way that it is not rational to make an arbitrary choice of what to believe. I also noted that desires are relevant to what to decide to do in a way in which they are not relevant to what to decide to believe. If one want something to be true, that can be a reason to try to make it true, but not by itself a reason to believe that it is true apart from what one does.

More needs to be said about how goals are relevant to practical reasoning. For example, one significant issue is whether there is a single category of goal, or a single measure of "utility", as opposed to a variety of functionally different things: desires, values, goals, intentions, commitments, principles, rules, etc. A related issue is whether we need to allow for a structure within goals in which some goals depend on others.

Here it is not enough simply to appeal to mathematical decision theory. In the simplest decision to which the theory applies (e.g. [von Neuman and Morgernstern, 1944]), one is faced with a decision between two or more exclusive acts, each of which has various possible outcomes to which one assigns certain values or "utilities". So $u(A)$ represents the utility of act A. One also assigns conditional probabilities, *prob(O,A)*, to each possible outcome O in relation to a given act A. Then the "expected gain" of a given outcome O of an act A is $\Sigma_A(prob(O,A) \times u(A))$.

The "expected utility" of each act A is the sum of the expected gains of each possible consequence of that act. Then the theory holds that rationality

requires doing the act with the highest expected utility or, if there is a tie for highest, one of the acts with highest expected utility.

The principles of mathematical decision theory are like principles of logic in being principles of consistency or coherence. So it is as much of a mistake to identify mathematical decision theory with the theory of practical rationality as it is to identify the theory of theoretical rationality with logic.

Some decision theorists argue that it is useful for individuals faced with hard practical problems to think of them in decision theoretic terms. Such individuals are advised to consider carefully what their possible acts are, what possible consequences each act might have, what utility they assign to each possible consequence, and how likely they think a given act would be to have a given consequence. They should then calculate expected utilities and choose the act with the highest calculated expected utility.

Is that good advice? The question can only be answered empirically. Do people do better using such a method or not? The suggested method is not obviously good advice. Given a poor enough assignment of utilities and probabilities, one can be led very wrong by one's calculation.

6.1 Derivative Goals

Some goals are derivative from others in a way that is important for practical rationality. One wants A. B is a means to A. So one wants B. That is, one wants B as a means to A. If one gets A in some other way, one no longer has the same reason to want B. Or, if one discovers that B is not going to lead to A, one no longer has the same reason to want B. It is irrational to continue to pursue an instrumental goal after the reason for wanting it has lapsed.

Also, consider the problem of deciding what to do when one has several goals. If one does A, one will satisfy goals $G1, G2$, and $G3$. If one does B, one will satisfy goals $G4, G5$, and $G6$. It is not easy to say how a rational person reaches an overall evaluation of acts A and B by combining his or her evaluation of the outcomes of each act. One idea (going back at least to [Franklin, 1817]) is to try to reduce the lists by matching outcomes of a with equivalent outcomes of B, cancelling the equivalent goals out, then considering only the remaining advantages of each course of action. That can still leave difficult choices.

But one thing can be said: one should not count the satisfaction of two goals as distinct advantages of an act if one's only reason for the first is that it will enable one to attain the second! That would be to count the same consideration twice.

Another point is that one cares about some things that are neither ultimate ends nor instrumental toward getting other things one wants. One can want good news, because that is evidence for something else one wants.

But the desire for good news is not a reason to try to influence what the news will be without influencing the event of which the news would be news.

6.2 Intentions

A rational person does not always reason directly from current goals, figuring out the best ways to maximize their satisfaction. That would resemble Special Foundationalism with respect to theoretical reasoning. It would ignore the role of long term intentions. Such intentions record decisions already made. Such decisions are not irrevocable, but they carry considerable weight and must not be frivolously discarded. A person incapable of maintaining long term intentions would be incapable of long term planning and would have only a low level of rationality [Bratman, 1987].

Intentions are not reducible to desires and beliefs, but put constraints on current planning of a special kind. A person's actual goals, as contrasted with things merely valued or desired, might be identified with what that person intends.

Intentions are directly related to action in ways not fully understood. Some authors think there are special intentions to do something *now*, where these are acts of will or volition, serving as immediate causes of action.

7 CONCLUSION

At present there is no mathematically elegant account of all aspects of reasoning. We have formal theories of implication and consistency, but these are only part of the subject. Logic, conceived as the theory of implication and consistency, and probability theory are not theories of reasoning. Conservatism, simplicity, and coherence are further features, with explanation, implication, and consistency being relevant to coherence.

BIBLIOGRAPHY

[Alston, 1989] W. Alston. *Epistemic Justification: Essays in the Theory of Knowledge.* Ithaca, NY: Cornell University Press, 1989.

[BonJour, 1992] L. BonJour. Problems of induction. In J. Dancy and E. Sosa (eds.), *A Companion to Epistemology*, pp. 391–5. Oxford: Basil Blackwell, 1992.

[Bratman, 1987] M. Bratman. *Intention, Plans, and Practical Reason.* Cambridge, MA: Harvard University Press, 1987.

[Chisholm, 1982] R. Chisholm. *The Foundations of Knowing.* Minneapolis, MN: University of Minnesota Press, 1982.

[Foley, 1987] R. Foley. *The Theory of Epistemic Rationality.* Cambridge, MA: Harvard University Press, 1987.

[Franklin, 1817] B. Franklin. *Private Correspondence, volume I.* London: Colburn, 1817.

[Goldman, 1986] A. Goldman. *Epistemology and Cognition.* Cambridge, MA: Harvard University Press, 1986.

[Goodman, 1965] N. Goodman. *Fact, Fiction, and Forecast*, 2nd edition. Indianapolis, IN: Bobs-Merrill, 1965.

[Hempel, 1965] C. G. Hempel. *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*. New York: Free Press, 1965.

[von Neuman and Morgernstern, 1944] J. Von Neumann and O. Morgernstern. *Theory of Games and Economic Behavior*. Princeton, NJ: Princeton University Press, 1944.

[Sober, 1988] E. Sober. *Reconstructing the Past*. Cambridge, MA: MIT Press, 1988.

[Stalnaker, 1984] R. Stalnaker. *Inquiry*. Cambridge, MA: MIT Press, 1984.

DAVID N. PERKINS

STANDARD LOGIC AS A MODEL OF REASONING: THE EMPIRICAL CRITIQUE

For over two millennia, since the days of Aristotle and Euclid, the notion of formal logic has figured centrally in conceptions of human reasoning, rationality, and adaptiveness. To be adaptive, the story goes, we must be rational about ends and means, truth and evidence. To be rational, we must reason about what means suit what ends, what evidence supports what conclusions. And to so reason, we must respect the canons of logic. It's common to note that transient moments of everyday cognition involve logical moves, at least implicitly. When you hear a dog bark outside, what you hear is a sound you recognize as a bark. You infer the presence of a dog, a deduction that might go:

> Around here, only dogs make the sound of a bark.
> I hear something that makes the sound of a bark.
> Therefore I hear a dog.

Examples like these intimate that formal logic is far more than a playground and workshop for philosophers, mathematicians, and designers of microchips. It is, if not the warp and woof of human reasoning, at least the warp, the woof perhaps being the beliefs from which we reason. Insofar as we are successful as a species in ways beyond the reach of chimpanzees, our logical prowess may be the cause.

Plausible as this story seems, considerable psychological research has profoundly challenged it over the past several years. Experiments have shown that people do not reason logically very well. Studies have suggested that when people appear to apply logic, they are really doing something else of a more constrained or situated character. Arguments based on an information processing perspective have urged that some of the demands of logical reasoning are unreasonable because unreachable—beyond the capacity of people or indeed any plausible sentient agent. In these and other ways, contemporary scholarship questions whether the logic of Aristotle and Euclid figures centrally in human cognition.

This array of findings inevitably has provoked a backlash, most notably by L. Jonathan Cohen, who, in a well known 1981 paper argued that this body of research had pretty much missed the point. Drawing on a distinction advanced by the linguist Noam Chomsky for the realm of language, Cohen argued that competence not performance was what counted. People might perform in erratic ways because of many circumstances—fatigue, lack of interest, deceptive aspects of a task. But underneath lies a strong competence to handle logical reasoning, a competence that would shine through under

supportive conditions. Cohen further argued that as a matter of principle one could not impugn human logical competence through empirical work, because the very standards for identifying shortfalls in performance derived from that competence.

Sampling the empirical literature, Cohen proposed that the attack on human rationality could be dissolved into four kinds of findings: (1) cognitive illusions, tasks that in one way or another prove momentarily deceptive because of quirks of human information processing; with enough prompted reflection, people would get it right; (2) tests of intelligence or education, tasks that involved specialized or sophisticated knowledge; if appropriately educated, people would get them right; (3) misapplications of appropriate normative theory, where people are measured by misapplied standards; (4) applications of inappropriate normative theory, where people are measured by mistaken standards.

So what to make of all this? Does the empirical picture compel a view of logical reasoning as poorly managed and error prone, not at all the centerpiece of human adaptation to a complex world? Or are the empirical results deceptive as Cohen argues? Or something in between or different altogether? This article reviews the evidence, appraises its weight, and offers a summative judgment of the place of logic in human thinking.

GROUND RULES

Before turning to the literature on this vexed issue, it's important to establish some ground rules for the inquiry.

First, what does logic mean here? It's convenient to speak of the topic as "standard logic," including the canons of formal deduction, the special case of disconfirming hypotheses by finding counterevidence for their implications, and also the principles of probabilistic and statistical inference developed by mathematicians over the past couple of hundred years.

Second, are we examining deliberate or reflexive reasoning? Stenning and Oaksford [1993] argue that standard logic or subsets of it can be implemented in quite different ways and that human cognition incorporates more than one implementation. The "bark-dog" inference and ones like it occur reflexively. Stenning and Oaksford aver that they show the operation of a rapid unconscious inferential system that lacks the full capacity of standard logic. In particular, it depends extensively on stored information about, for instance, barks and dogs; it cannot accommodate very many novel entities (nixelmugs, brazzbugs, feegleipips) or contrafactual entities (cup means telephone, telephone means saltshaker, saltshaker means tiger, etc.) in an episode of reasoning. In contrast, the kind of deliberate reasoning people do with syllogisms or like forms at least in principle allows unfamiliar and contrafactual entities in any number, granted that we may get confused or

exceed cognitive capacity and need the help of pencil and paper.

Almost all the research on the role of standard logic in human thinking concerns deliberate rather than reflexive reasoning. Accordingly, the present analysis will focus on deliberate reasoning and the place of standard logic in it.

Third, recalling Cohen's [1981] concern, are we examining performance or competence? This analysis will emphasize performance. As several critics of Cohen have pointed out (see the discussion section of [Cohen, 1981]), in practical adaptive circumstances it's performance that counts. However, the issue of competence will resurface from time to time.

Fourth, we need to be clear about what standard logic might aspire to explain. At best standard logic proposes necessary conditions for reasoning. There is no question of sufficient conditions because, as a number of authors have emphasized, standard logic has no sense of direction. It says nothing about what inferences are worth making, what propositions are worth defending for practical or theoretical ends. Standard logic cares not whether we infer the Pythagorean theorem from Euclid's axioms or the unmarried state of bachelors. Standard logic establishes boundaries within which reasoning can properly proceed, but gives no compass for navigating within those boundaries.

With this limit acknowledged, consider an aspiration within the conceivable reach of standard logic, a bold and idealistic claim about standard logic as a model of human reasoning:

> THE AMBITIOUS CLAIM: *By and large, people can, should (in the sense of adaptation), and do reason according to standard logic.*

Many would feel that the Ambitious Claim claims far too much. In fact, it is something of a straw man, designed from the first to be hewed down. But even as a straw man it proves remarkably useful. Its three subclaims of "can," "should," and "do" provide an analytical framework for examining the debate. Starting with the "do," let us examine whether or to what extent people do reason according to standard logic. That leads into an analysis of whether people can, and then whether they should.

DO PEOPLE REASON ACCORDING TO STANDARD LOGIC?

One of the simplest questions to ask about standard logic and human reasoning is whether people get the results, and only the results, that standard logic sanctions. For now, the issue is not whether people get the results through manipulating a rule system substantially the same as standard logic but rather whether, whatever the mechanism at work, standard logic describes the answers people give.

For a number of years, psychologists have examined human reasoning on a wide range of logical tasks. These include, for instance, syllogistic reasoning tasks in abstract and concrete forms, tasks that ask what evidence people would need to test an implication, tasks involving the law of large numbers and questions of sufficient sampling, and tasks involving combining probabilities, as in conjunctive probabilities or Bayes' theorem.

There is no need here to review the full range of this research. Good general sources include, for example, [Evans *et al.*, 1993; Flamagne, 1975; Kahneman *et al.*, 1982; Nisbett, 1993; Nisbett and Ross, 1980; Revlin and Mayer, 1978]. A surprisingly rich microcosm of the pattern of findings emerges from considering two very simple inferences, modus ponens and modus tollens.

How Well People Perform

Modus ponens involves inferences of the form If A then B; A; therefore B. For a concrete example, if the candy is blue, it's a mint; the candy is blue; therefore it's a mint. Adults reason quite reliably in accordance with modus ponens, asserting such inferences or affirming their validity. The most typical experimental paradigm presents subjects with a candidate inference and asks them to affirm or deny its validity, or presents alternative inferences including "nothing follows." In a review of the extensive research on deductive reasoning, Evans, Newstead, and Byrne [1993, p. 36] tabulated responses on modus ponens as well as other inferences from a range of studies. The studies focused on conditionals that asserted abstract or arbitrary relationships. For instance, in the above example, there is no reason to expect blue candies to be mint or not to be mint. Endorsements of modus ponens inferences ranged from 89% to 100% over 11 experiments, with an average of 97% .

Modus tollens is a different story. Modus tollens involves inferences of the form If A then B; not B; therefore not A. If the candy is blue, it's a mint; the candy is not a mint; therefore it's not blue. Evans *et al.* [1993, p. 36] report that this inference proves considerably more challenging. Over the same range of studies mentioned above, people endorsed modus tollens inferences from 41% to 81% of the time.

People's reasoning depends not only on the basic forms of inferences but on negations in either A or B, which complicate the processing required. Evans *et al.* [1993, p. 43] report several studies exploring this. Modus ponens survives the introduction of negations largely unscathed, with rates of endorsement ranging from 90 to 100% . However, modus tollens proven particularly sensitive to the variation: If not A, then B; not B; therefore A: If the candy is not yellow, then it's sour; it's not sour, therefore it's yellow. People find inferences in this form difficult, endorsing them from 12% to 34% of the time.

Modus tollens commonly catches people not endorsing a conditional inference that they should. Of course, people also commonly endorse conditional inferences that they should not. This takes the form of two classic errors of reasoning, affirming the consequent and denying the antecedent. Given If the candy is blue, it's a mint, affirming the consequent takes the form: The candy is a mint; therefore it's blue. Denying the antecedent takes the form: The candy is not blue; therefore it's not a mint.

Interpreting people's performance presents challenges, because often people treat the conditional as a biconditional. Contextual factors can encourage people to do so, and thus create the appearance of affirming the consequent or denying the antecedent when people are extrapolating from the information given in what they take to be an entirely reasonable way. The candy example is a case in point. The candy manufacturer could have color-coded the candies, so the blue candies and *only* the blue candies are mints. This need not be so, but it might well be so, inviting the presumption of a biconditional. There are many experimental demonstrations of such influences.

However, when people deal with conditionals where no such factors appear to be in play, they still often affirm the consequent or deny the antecedent. In the Evans *et al.* [1993, p. 26] review, rates of affirming the consequent ranged from 23 to 75 percent; rates of denying the antecedent ranged from 17 to 73 percent. One interpretation might be that many subjects routinely take the conditional as a biconditional, even though these studies used content that did not invite this. However, experimenters have investigated this possibility subject by subject. Evans *et al.* report that only about 50% of subjects show a consistent pattern, either conditional or biconditional.

Failure to endorse correct inferences and the endorsement of incorrect inferences are phenomena widely documented across investigations of many kinds of reasoning. To offer a different kind of example, people often neglect statistical factors altogether in their reasoning. In one classic example, the psychologists Daniel Kahneman and Amos Tversky [1982] posed this problem:

A certain town is served by two hospitals. In the larger hospital about 45 babies are born each day, and in the smaller hospital about 15 babies are born each day. As you know, about 50% of all babies are boys. The exact percentage of baby boys, however, varies from day to day. Sometimes it may be higher than 50% , sometimes lower.

> For a period of 1 year, each hospital recorded the days on which (more/less) than 60% of the babies born were boys. Which hospital do you think recorded more such days?

Most subjects thought that the two hospitals probably recorded about the same number of days, and about equal numbers gave the edge to one

hospital or the other. However, statistically, the hospital with 15 babies per day is much more likely to exceed 60 percent males. A 10 percent deviation from the mean becomes increasingly improbable with a larger sample.

If this task seems to expect too much of subjects' understanding of the law of large numbers, a different kind of problem reveals a more blatant failure in probabilistic reasoning. A cornerstone of probability theory says that the conjunction of two events can be no more probable than either alone. Yet people often violate this condition. In one study, for example, subjects read the following description [Tversky and Kahneman, 1982]:

> Linda is 31 years old, single, outspoken, and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations.

Then they judged which of two descriptions was more probable:

(a) Linda is a bank teller.

(b) Linda is a bank teller and is active in the feminist movement.

Eight-seven percent of the subjects chose (b) over (a) as more probable. A series of studies by Tversky and Kahneman have confirmed this effect over and over. Subjects tend to use what the authors call a "representativeness heuristic." The subjects make probabilistic judgments based on what sounds like a representative description, neglecting principles of probability.

In another demonstration relevant to representativeness, Kahneman and Tversky [1982] told one group of subjects that 30 engineers and 70 lawyers had been interviewed and personality profiles extracted. The subjects read five of these profiles and predicted whether each person profiled was more likely one of the engineers or one of the lawyers. Another group undertook the same task, except they were told that 70 engineers and 30 lawyers had been interviewed. The manipulation changes the *a priori* probability for a person being an engineer or a lawyer. However, this made absolutely no difference to the subjects. They were as likely to judge a profile to be that of an engineer in both conditions.

These and numerous other studies addressing many forms of inference warrant a general conclusion that contradicts the Ambitious Claim: In fact people commonly do not reason in accordance with standard logic. They do not endorse what logically follows, or do endorse what does not follow.

Can We Explain Away the Shortfalls?

It's natural to wonder whether such lapses in standard logic really impugn people's ability to reason logically. Several defenses merit consideration.

Perhaps the simplest is occasional careless error. However, the rates of deviation from standard logic are often greater than 50 percent. Accordingly, the deviations can hardly be interpreted as occasional careless error.

Another defense points out that people are often reasoning reasonably in the basis of background knowledge and the pragmatics of discourse (e.g. [Grice, 1975]). When you are told, for example, "If you answer the question correctly, you get the prize", it's reasonable to conclude that if you don't answer the question you will not get the prize. People rarely use the explicit logical locution "if and only if" outside of certain academic contexts. They don't need to. Instead, the pragmatics of communication and human interaction make the biconditional plain.

However, experimenters have systematically manipulated such factors to appraise their impact. The trend of their influence is much as one would expect, but in tasks that minimize their influence, for instance the conditionals used in the studies mentioned above, large deviations from standard logic still occur.

Another defense of people's reasoning might hold that they reason perfectly well with the information as encoded. Henle [1978], a well-known researcher on reasoning, averred, "I have never found errors which could unambiguously be attributed to faulty reasoning." Henle noted that subjects instead might fail to accept the logical task, misstate premises or conclusions, and so on. But such a defense of human reasoning also has its problems. First of all, while encoding premises as stronger than they are may account for affirming the consequent and denying the antecedent, it's harder to see how one could encode an if-then premise so that a modus tollens inference would not follow. More generally, although many seeming errors in inference may reflect faulty encoding of input or faulty articulation of output, from a practical adaptive standpoint reasoning performance necessarily involves all three stages: encoding, inference, and output of results.

For another defense, Cohen [1981] would see the kinds of shortfalls discussed here as falling in his categories (1) or (2), matters of rectifiable cognitive illusion or lack of pertinent technical knowledge. At least for the shortfalls discussed here, he would seem to be right. The logic of modus tollens can be made plain readily enough. With some prompting, people readily recognize that their reasoning on the "Linda" problem must be faulty [Tversky and Kahneman, 1982]. Nisbett [1993] offers a number of studies that show that people can learn important aspects of probabilistic and statistical reasoning and perform better on the kinds of tasks psychologists have used in studies of poor reasoning.

However, the point remains that human *performance* is commonly shoddy, whatever the underlying competence. Therefore, although such findings may not threaten fundamental human rationality, they certainly raise questions about the functionality of standard logic as an adaptive strategy in day-to-day affairs.

A Question of Mechanism

Despite the negative findings, people are far from indifferent to the constraints and permission of standard logic. Evans, Newstead, and Byrne [1993] point out that people almost always perform considerably above chance on logic problems. Even when the problems are quite abstract, offering no obvious anchor in experience, people show a measure of logical ability. Certainly to some extent, people reason in accordance with standard logic.

But what does "in accordance" mean? On the surface, it simply asks that people endorse those and only those inferences warranted by standard logic, as well as generating a good many of those inferences themselves (if people could only endorse but not generate, their capacity would only serve to test already proposed inferences rather than to discover conclusions). But underneath, in terms of psychological process, there is more than one way to look at how people reason.

The most obvious theory mirrors how a logician might think about logic: People have rules represented somehow in their minds. Faced with a logic problem, people deploy those rules to generate inferences or test given inferences. This becomes more difficult and errors creep in as the tasks require more inferential steps.

Several investigators have advocated rule theories (e.g. [Braine and O'Brien, 1991; Rips, 1990; Rips, 1994]. While there is room for some variation, the basics are rather similar. A rule theory might account for modus ponens as follows:

1. . You are informed, "If the candy is blue, then it's a mint" and "The candy is blue."
2. You discern the logical form of the givens: $P \to Q$, and P.
3. You retrieve a rule that happens to fit perfectly, some representation of modus ponens.
4. Applying the rule, you conclude, "The candy is a mint".

This inference proves easy because you happen to have a rule that exactly matches its need. Given the demonstrated difficulty of modus tollens, people lacking formal training apparently have no such rule. Thus, modus tollens inferences require a chain of rule applications, perhaps something like this:

1. You are informed, "If the candy is blue, then it's a mint" and "It's not a mint".
2. You discern the logical form of the given: $P \to Q$, and Not Q.
3. You have no rule you can apply directly. You may judge mistakenly that no conclusion is possible, as many people do.

4. Alternatively, your rule system includes, let us say, a reductio ad absurdum rule. From this and other rules, you can construct a proof that the candy is not blue: Suppose the candy is blue; then it's a mint (by modus ponens); but it's not a mint, a contradiction; therefore (by the reductio) Not the candy is blue.

5. So you conclude, "The candy is not blue".

Johnson-Laird [1983] proposed a rather different conception of logical reasoning in his *Mental Models*. Johnson-Laird and Byrne [1991] provided a well-worked out theory of propositional reasoning according to mental models. Model theory says that people reason not by the syntactic manipulation of logical forms, as rule theory holds, but by mentally representing situations that might occur, excluding situations eliminated by premises.

For example, when you are told, "If the candy is blue, then it's a mint," model theory says that you might envision three situations:

A blue candy, a mint
A nonblue candy, a mint
A nonblue candy, not a mint

These three situations exhaust the possibilities consistent with the given conditional. However, model theory recognizes that constructing complex representations burdens human short-term memory. Consequently, rather than elaborating all possible situations, a person would normally construct an abbreviated model along these lines:

A blue candy (all blue candy situations appear here), a mint
... other situations can be generated ...

The first line says that all situations involving a blue candy appear on this line. The combinations not involving a blue candy are all allowed and can be generated if needed.

The inference process proposed for inferences like modus ponens and modus tollens is quite different from that of rule theories. Modus ponens occurs as follows:

1. You are informed, "If the candy is blue, then it's a mint" and "The candy is blue".

2. You construct an abbreviated model of the first premise:

   A blue candy (all blue candy situations appear here), a mint.
   ... other situations can be generated ...

3. And a model of the second premise:

   A blue candy.

4. You try to combine the two into a single model. This is easy since the situation described under (3) directly matches the situation described under (2) and eliminates the tacit nonblue situations under (2). There results a one-situation model:

   A blue candy, a mint.

5. So you conclude, "The candy is a mint".

In contrast, a modus tollens inference is much more complex, in keeping with its less frequent endorsement:

1. You are informed, "If the candy is blue, then it's a mint" and "The candy is not a mint".

2. You construct an abbreviated model of the first premise:

   A blue candy (all blue candy situations appear here), a mint.
   ... other situations can be generated ...

3. And a model of the second premise:

   Not a mint.

4. You try to combine the two into a single model. But *not a mint* does not match anything in your abbreviated model of the first premise, so you cannot proceed straightforwardly. You may conclude that no inference is possible, as many people do.

5. Alternatively, you may expand (2) into a full model:

   A blue candy, a mint.
   A nonblue candy, a mint.
   A nonblue candy, not a mint.

6. You can combine this model with *Not a mint.* The only situation consistent with both is the one described in the third row, yielding the merged model:

   A nonblue candy, not a mint.

7. So you conclude, "The candy is not blue".

An ample literature shows that both rule theories and model theory account for quite a range of experimental results, including for example the reliability of modus ponens and the difficulty of modus tollens. But which describes human reasoning better?

One objection to model theory proves misplaced: Superficially model theory asks people to form mental images of situations. But some scholars

doubt the very existence of mental images while others may not report any mental images while grappling with such problems. However, Johnson-Laird and Byrne [1991] make it clear that model theory requires no such assumption. The representations of possible situations can be highly schematic. They can take the form of tokens or tags rather than imagined scenes. In fact, the form of representation is irrelevant to model theory. What counts is the structure: the itemization of alternative situations in a fully expanded model, along with abbreviated forms for cognitive parsimony.

Some investigators argue that model theory has an edge in explaining conditional reasoning behavior. It appears to predict and account for findings not readily accommodated by rule theories. One case in point is the relative difficulty of modus tollens for the conditional and the biconditional. Rule theories predict that even with a biconditional premise ("If and only if the candy is blue, it's a mint"), modus tollens is a very indirect inference, requiring a reductio ad absurdum maneuver as described above. In contrast, model theory predicts a somewhat more direct modus tollens inference because the model for a biconditional requires less expansion to get to the negative cases. Your abbreviated model of "If and only if the candy is blue, it's a mint" looks like this:

> *A blue candy (all blue candy situations appear here), a mint (all mint situations appear here).*
> *. . . other situations can be generated . . . .*

And your expanded model looks like this, with only two situations rather than the three that result when you expand the model for "If the candy is blue, it's a mint."

> *A blue candy, a mint.*
> *A nonblue candy, not a mint.*

So expanding a model for the first premise is easier and merging that with the model of the second premise is easier, predicting an easier modus tollens for the biconditional than the conditional. In keeping with this, Johnson-Laird, Byrne, and Schaeken [1992] showed that subjects endorsed modus tollens substantially more often for biconditionals than for conditionals.

However, this is a tricky issue in two respects: what rules a competing rule-based system has at its disposal, and what counts as a rule-based system. Regarding the first, Rips [1994] offers a theory and computer simulation of rule-based reasoning, including rules for backward inference. Rips' model accounts for an impressive range of findings. Regarding the second, recall that model theory as elaborated by Johnson-Laird and Byrne [1991] involved models that could be quite token-like. A model-based modus ponens or modus tollens carried out with a tokenized representation could be seen as a process of rule application. The rules would not be inferential rules

like modus ponens but rules nonetheless, ones about merging tokens. Rips [1994] makes very much this point, suggesting that, when the schematic character of what Johnson-Laird and Byrne view as models is recognized, the contrast between rule-based and model-based accounts appears minor.

Assuming a sound contrast between rule-based and model-based accounts, other qualifications are in order. One is that people may have multiple methods of reasoning. Perhaps modus ponens is handled by direct application of a rule even if some other inferences call upon models. As mentioned earlier, standard logic can be implemented in more than one way (for instance by rules or by models) and it's quite possible that deliberate human reasoning does just that [Stenning and Oaksford, 1993].

For a second qualification, whatever the case with conditional reasoning, it's clear that for some kinds of reasoning some people reason with mental models while others use rules. Roberts [1993] reviews several such findings. One example is the linear syllogism. This refers to problems such as "Alfredo is taller than Betty; Clarence is shorter than Betty; can you infer anything about Clarence and Alfredo?" You can of course conclude that Clarence is shorter than Alfredo. One way to do so is through application of transitivity, converting "Alfredo is taller than Betty" to "Betty is shorter than Alfredo." Another way is to sketch or imagine side-by-side vertical lines standing for Alfredo, Betty, and Clarence. You construct the diagram based on the premises and simply read off the relation between Alfredo and Clarence. Some people do it one way, some another, and some may use mixed strategies.

For a third qualification, whatever people usually do, there is no doubt that people can learn to apply a rule-like approach fluently. If I may use myself as an example, my academic background is mathematics. When I look at most of the problems posed to subjects in such experiments, I reflexively deploy rules. When I see "not a mint" I immediately apply a modus tollens. When I see a negation like not (red and sweet), I immediately translate it into not red or not sweet. This says nothing about what untrained people usually do. But it does say that there is nothing psychologically impossible about rule-based reasoning.

One possible approach to resolving the debate between model theory and rule-based approaches is to say that people have an inherent rule-based or model-based (depending on your position) competence that sometimes gets overlaid by learned or invented strategies of the opposite kind. However, Roberts [1993] argues that this approach lacks genuine leverage: One would need to specify a criterion for deciding across a range of different reasoning tasks which is the strategy—rule-based or model-based reasoning—and which the underlying competence. No such criteria have been proposed, nor does it seem easy to formulate criteria that accommodate the current range of findings without begging the question. Roberts urges that the debate has been marked by a tendency not to face up to the complexity of the phenom-

ena and the challenge of individual differences. As he somewhat ironically puts it:

> A genuinely parsimonious theory is *the simplest possible theory that accounts for the data.* In this context, the performance of individuals (or tasks) who are out of line is being ignored, as this would require a more complicated theory, and hence the principle of *pseudo*-parsimony would be a more appropriate description for such a simplification, this being *the theory that accounts for the simplest possible data* [Roberts, 1993, p. 578] (italics Roberts').

The issue becomes all the more confusing due to differences in what rule-based reasoning might mean. Smith, Langston, and Nisbett [1993] systematically analyze the case for rules for reasoning. They offer eight criteria that suggest a person is reasoning with rules. Here are five that seem particularly suggestive (pp. 367–8):

- Performance on rule-governed items is as accurate with unfamiliar as with familiar material.
- Performance on rule-governed items is as accurate with abstract as with concrete material.
- A rule, or components of it, may be mentioned in a verbal protocol.
- Performance on a specific rule-governed problem is improved by training on abstract versions of the rule.
- Performance on problems in a particular domain is improved as much by training on problems outside the domain as on problems within it, as long as the problems are based on the same rule.

Surveying a wide range of studies with these criteria in mind, Smith et al. conclude that several kinds of inferences live up to the standards for rule-based reasoning: modus ponens but not modus tollens; permissions and obligations; causal rules involving sufficient and necessary, just sufficient, and just necessary causes; and the law of large numbers. Note here that people often do not respect the law of large numbers, as discussed earlier. However, Nisbett, Krantz, Jepson, and Kunda [1993] present evidence that they do in straightforward problems that involve comparative judgments of the reliability of larger and smaller samples. Also, Nisbett [1993] offers a variety of evidence that, however they start out, people can learn to apply statistical principles in a rule-like fashion across diverse domains.

The argument of Smith, *et al.* [1993] may seem to oppose model theory as articulated by Johnson-Laird and Byrne [1991]. But in fact it does not,

as Smith, *et al.* [1993, p. 390] seem to agree. As noted earlier, the reasoning process proposed by Johnson-Laird and Byrne involves rules operating on highly tokenized representations. Johnson-Laird and Byrne might say, "Yes, Smith et al.'s criteria show that rules for reasoning are in play; but the rule may not always be what it seems—what looks like application of a modus ponens may in fact be the application of a rule for merging models that yields the same effect."

What can be said in summary about the debate concerning rule-based and model-based theories of reasoning? If we take rule-based theories of reasoning to mean reasoning with any rules whatsoever, with Smith *et al.* [1993] criteria in mind, then at least tokenized versions of model-based reasoning become special cases of rule-based reasoning. The question "which do people really use?" is ill-constructed.

In contrast, if we take rule-based theories of reasoning to mean theories proposing that people reason with rules of the sort found in logic texts, then some empirical research today argues that aspects of conditional reasoning are model-based. However, empirical research does not justify the conclusion that all deductive reasoning is model based. On the contrary, the style of reasoning varies across tasks and across individuals within tasks, people adopting different strategies reflecting largely unstudied task and individual variables. Rips [1994] argues that at least a limited architectural core of reasoning involving AND, OR, NOT, and IF is rule-based, whatever may be the case for various strategic and pragmatic extensions of this core.

At the very least, the debate between the rule and the model approach shows that reasoning "according to" standard logic can mean more than one thing. The constraints and permissions of standard logic can be realized through more than one mechanism. People in effect capitalize on this and adopt or devise reasoning strategies to suit the occasion.

If Not, Why Not?

The track record of human performance shows clearly that people often do not reason correctly according to standard logic. This directly challenges the third term of the Ambitious Claim.

But more than that, indirectly it raises questions about the "can" and the "should" clauses of the claim, as well. In fact, the puzzle comes from an application of modus tollens: If indeed people can and should reason according to standard logic—if people are capable of it and it is functional to do so—then people as adaptive organisms could be expected to do so, at least most of the time. But people, it seems, do not. So by modus tollens the premise must be false. Perhaps people cannot reason according to standard logic: They run afoul of cognitive limits of one sort or another. Or perhaps they should not: Standard logic looks to be more adaptive than it is, other seemingly more limited reasoning strategies serving us better.

Or both of these.

In fact, the literature offers challenges both to the "can" and the "should." Let us consider each category in turn.

CAN PEOPLE REASON ACCORDING TO STANDARD LOGIC?

At first blush, the question of whether people can reason according to standard logic seems odd. With modus ponens and modus tollens in mind, the inevitable answer seems to be, "Of course they can. They may not, but surely they can. It's just not that complicated."

Indeed, the most basic moves of logic are not that complex. If people do not make them reliably, surely most people can learn to do so, at least when facing explicit problems. The basics of logic are considerably easier than the basics of arithmetic.

The catch, of course, is that such elementary applications of standard logic do not get to the heart of the issue. The question is more whether people generally can manage the more complex patterns of logic often needed to get their beliefs, plans, and decisions straight and get along better in the world. Isolated deployments of modus ponens, modus tollens, and whatnot will hardly build a life. Can people handle the more intricate maneuvers that might?

Limits of Computation

One kind of argument looks to limits on human computational capacity. While one modus ponens or modus tollens strains few brains, what about situations that demand several inferences? It's well known that people have difficulty holding in mind more than a few items at a time [Miller, 1956], and thus utterly unsurprising that complex problems baffle people, especially when they lack memory supports like pencil and paper. Evans *et al.* [1993] identify complexity as one of the key variables that causes lapses in formal reasoning.

Moreover, how much complexity must we handle to conduct ourselves in a rational way? Interestingly, certain natural ideals of rationality pose computational demands that outstrip not only the ready reach of human beings but also that of any sentient entity.

Realizing potential knowledge. One such ideal is to come to know explicitly what we potentially know—what's implied by the beliefs we have. As Plato urged in his *Meno,* having Socrates lead Meno's slave boy to recognize that a square erected on a square's diagonal has twice its area, potentially we know deep and subtle features of the world that may have utterly eluded us so far. Such potential knowledge reaches beyond mathematical insights to include the utterly prosaic as well: the ingredient you

should have added to the cake if only you had drawn the implication from the kitchen disaster story your neighbor told you last week, or what tax loophole you might legitimately duck through.

The trouble with this ideal is that it is computationally unwieldy. One cannot infer everything. The infinity of possible entailments makes too much work for any conceivable mind. That aside, chasing chains of inference outward toward their remote ends makes error inevitable: Somewhere an invalid step would occur, making the rest of the journey utterly misleading. And that aside, possible entailments spread out in all directions. Which way to go? As noted earlier, standard logic has no compass.

Checking for consistency. If realizing potential knowledge poses an unreasonable ideal, a more tractable aspiration for the rational mind might be simple consistency. If we cannot search out the remote implications of our beliefs-so-far, at least we might ensure that those we have do not war with one another. We might hope to distribute praise or blame by constant criteria, exercise the same critical standards for astronomy and astrology, believe in our adult children's judgment or not so much or for some things and not others but not erratically and whimsically, and so on.

Modest and sensible as this ideal sounds, it too proves to be computationally intractable. Cherniak [1986, Chapter 4] examines the demands of checking for consistency, demonstrating that the computation required expands exponentially with the size of the database of beliefs. Consistency checking sounds nice, but it's too much work.

Revising beliefs. If we cannot realize all our potential knowledge or keep a thoroughly consistent database of beliefs, at least we might aspire only to believe what we have good reason to believe—what follows from whatever general beliefs and particular lessons from experience we deem to be foundational. By this criterion, we should prune out whatever lacks such grounds. When we get new information impugning beliefs we previously thought to be adequately grounded, we should demote them from "true" to "plausible" or "doubtful".

Experimental studies show that people often do not behave in this way. For instance, suppose you are asked to persuade people to donate blood. You develop a positive or negative impression of your persuasive ability, based on people's responsiveness. Later, you are told that the whole experience was a fabrication, the people you hoped to persuade actors. What happens to your impression of yourself as a good or not-so-good persuader? Although its basis has been totally discredited, the impression lingers. This finding by Jennings, Lepper, and Ross is summarized along with several others to the same effect in Ross and Anderson [1982].

Unfortunately, belief revision too is computationally unwieldy. Harman [1986] argues that the beliefs we hold often have complex histories, deriving from a myriad of sources. It's unmanageable for us to keep track of the history of justification behind the beliefs we hold. Consequently, when

something impugns the original basis for one of our beliefs without directly casting doubt on it, we are not typically in a good position to revise it.

In summary, certain ideal aspirations associated with standard logic set unreasonably high standards for the human mind. We creatures cannot live up to such standards, even if we sought to commit ourselves to the enterprise. Nor is this just a human frailty. Any cognitive entity—for instance, an artificially intelligent machine—would face similar challenges.

This does not mean that people should or do proceed with blithe disregard for the implications of what they know, inconsistencies in their beliefs, or occasions where they might revise their beliefs. People commonly attend in a partial way to such matters. Harman [1986] argues that a normative policy for a limited cognitive agent is maintaining coherence. You only challenge a belief you hold if reasons emerge to doubt it. You revise your beliefs minimally, responding to conflicts you encounter or capitalizing on occasions to introduce stronger patterns of coherence. One might put it this way: ideals like those above are best pursued opportunistically, when circumstances present problems or invitations. Making much more of a mission of it than that is not just maladaptive but impossible for the organism.

Limits of Intelligence

Another kind of limit worth exploring concerns the variation from person to person in the ability to handle the demands of standard logic. Faced with a complex problem of inference—for instance, one of those puzzles that begins "Alice is the sister of Roger; Roger lives next door to his half-brother" and so on, some people much more so than others will prove able to manage the intricacies and extract a valid inference like "Roger forged his grandmother's signature on the will."

The most obvious measure of this capacity is IQ. Advocates of IQ see it as a general measure of people's ability to handle complex novel intellectually challenging tasks of any sort (e.g. [Herrnstein and Murray, 1994; Jensen, 1980]). A huge body of research documents the predictive relationship between IQ and a wide range of intellectually demanding performances, including school learning and job performance (e.g. [Brody, 1992]). The relationship is far from perfect—correlation coefficients around .5 are typical, accounting for 25 percent of the variance—but it is quite real and robust.

IQ is a controversial topic, in several respects. How well does IQ reflect intelligent functioning in the world? To what extent is IQ genetically determined? Can one gain a higher IQ through the right kinds of study? Is IQ only one kind of intelligence among several? The challenges to the centrality of IQ have been many (e.g. [Ceci, 1990; Gardner, 1983; Gould, 1981; Perkins, 1995]). To explore these topics seriously would take us far afield. For present purposes, let me suggest two points: First, IQ is a limited construct with less power to explain real-life intelligent function-

ing than has often been thought, as argued in the sources just mentioned. But second, IQ nonetheless measures variations in the capacity of people to process complex novel information (e.g. [Brody, 1992; Perkins, 1995; Jensen, 1980]).

Let us return to the topic of standard logic. Certainly the exercise of standard logic demands both coping with complexity and dealing with novelty. Many of the inferences demanded not just by puzzles but also by everyday circumstances, if cast into logical form, pose considerable complexity. As to novelty, the whole point of standard logic is its content-free character. The machinery of inference should apply equally to any topic, however familiar or unfamiliar, concrete or abstract. Indeed, it's worth recalling that this was one of the criteria Smith *et al.* [1993] advanced for recognizing occasions when people reason with rules. Both these points suggest that people's capacity to deploy standard logic will vary considerably with IQ or like measures.

In keeping with this, many items used on IQ tests have a logical character. For instance, test-takers commonly face number series problems like: 1, 4, 9, 16, with an invitation to select the next term from several options. Raven's Matrices, a supposedly culture-free test of intelligence, presents test-takers with a three-by-three grid of somewhat similar diagrams. The diagrams vary in detail in a systematic way as one reads across, and in another systematic way as one reads down. The bottom right box is blank. The test-taker tries to infer the two rules in play and select from a set of options the diagram that would complete the pattern. Verbal analogy problems (feathers are to hen as wool is to—a. blankets, b. sheep, c. soft, d. animal) ask test-takers to discern the principle governing the first analogy and apply it to complete the second.

Note that such items are not strictly deductive. The given information does not usual entail an answer. For instance, the number series 1, 4, 9, 16 invites interpretation as the squares of the positive integers, but many other polynomials can be written that yield those first four terms with different continuations. However, as a matter of practicality, items are designed so that rival interpretations consistent with the given information occur rarely. In effect, the problems become challenges of hypothetico-deductive inference—discerning possible patterns and checking your hypotheses against the given information.

In summary, this line of evidence argues against the notion that people can reason uniformly well according to standard logic. Here one might well ask, "So what?" The claim that people reason according to standard logic hardly demands that everyone reason equally well.

However, there is something at issue here. People's grammatical performance offers an illuminating analogy. As has often been pointed out, almost everyone masters the grammatical structures of their native language to a high degree of sophistication. Despite many intricate constructions

and innumerable exceptions, people achieve effortless fluency and accuracy. Seeming shortfalls in a person's grammar typically reflect not some lack of mastery but the idiosyncratic local grammar of the person's upbringing. Thus, grammatical performance sets a kind of norm for what it might mean for people to behave in accord with a set of core principles.

Plainly, people's reasoning gauged by standard logic falls far short of this norm. Not only do people often diverge from standard logic, but also their ability to deal with complex instances of standard logic varies widely. This challenges the notion that standard logic plays a central role in organizing reasoning, as grammar does in organizing our verbal communications. If it did, we should be better at it, as we are with grammar. Bearing in mind the intent behind the Ambitious Claim that people can, should, and do reason according to standard logic, the claim should be interpreted with some strictness and substantial individual variations challenge it.

Limits of Understanding

Another constraint on how well people reason according to standard logic is understanding. Often people simply misunderstand the logical demands of the task. In some sense, this holds true whenever people diverge from standard logic. When, for instance, a person fails to affirm a modus tollens, the person fails to understand the necessity the modus tollens reflects. However, a stronger version of limits of understanding recognizes that sometimes people do not just miss the point but maintain sharply contrary conceptions of the logic of the situation.

For example, a study by Byrne and Handley [1992] involved subjects interpreting statements of the form NOT (A AND B) and NOT (A OR B). Subjects generally misinterpreted these forms, 75 percent and 58 percent of the time respectively. The most common error treated both as meaning NOT A AND NOT B.

Chazen [1989] examined high school students' understanding of entailment in geometry as part of an investigation of their learning in that subject matter. He discovered a provocative paradox in their conception of proof, even after considerable instruction. He asked students to suppose they had proved something. Later on, might they find an exception? Many students responded that yes, by looking hard one might. He also asked them whether they felt a statement must be true for sure after they saw three or four instances of it. Many said yes.

These students handled the mechanics of geometry well enough. They could produce proofs. But they had somehow missed the whole point of the enterprise: Examples do not suffice, since a theorem might not hold for the next example; but formal deductive proof establishes the theorem for all cases. Instead, many held the reverse conception: A formal proof establishes the theorem for most but maybe not all cases; but a few cases

establish the theorem absolutely.

To reach beyond the area of deduction, probabilistic reasoning is a veritable swamp of misconceptions. One of the best known is the "gambler's fallacy", the belief that over a series of trials chance events somehow seek out the average, correcting prior divergences. For instance, suppose you flip a coin five times and each time the coin happens to come up heads. Subscribers to the gambler's fallacy believe that tails are "due". On the next flip, the coin has a better than 50-50 chance of coming up tails [Lindman and Edwards, 1961].

In summary, another factor limiting people's reasoning in accordance with standard logic is their understanding. It's not just that people fail to recognize valid entailments; they often exercise mistaken principles and conclude that something follows which does not. Very likely they could learn better principles, but they have not.

Here one might wonder how this really challenges the Ambitious Claim. We don't complain when people fail to understand carburetors or counterpoint without an opportunity for special training. Why should we treat such lapses as these as limits on people's reasoning according to standard logic? Cohen [1981] would want to dismiss many of these findings as cognitive illusions (his category 1) or consequences of ignorance (category 2). And well he might, in support of his competence defense of human rationality. However, as in the previous section, behind the Ambitious Claim lies the idea that standard logic is a core performance strength for human beings. So people ought to develop the ability spontaneously, as with crawling and walking. Or people ought to learn it from the interactions with and around them, as with grammar. But this does not happen.

Does the "Cannot" Explain the "Do Not?"

In summary, these arguments challenge the "can" clause of the Ambitious Claim. People cannot so readily reason according to standard logic, because of limits of computation, limits of intelligence, and limits of understanding. Standard logic may be a good way to reason in principle and a good candidate for a core competence in human cognition, but it poses an unapproachable ideal in some respects. Moreover, even setting computational limits aside, human reasoning as measured by standard logic lacks the marker of a core competence such as grammar: the reliable mastery with little individual variation across a wide range of contexts. It may well be that some small subset of standard logic involving modus ponens and a few other kinds of inference meets this criterion [Rips, 1994], but certainly not the full compass of standard logic.

It's tempting here to conclude that the failure of the "can" clause explains the failure of the "do" clause: People do not reliably reason according to standard logic simply because they cannot. However, this overextends the

argument. Although the limits discussed above seem quite real, they do not account for a number of cases where people diverge from standard logic.

Recall again the case of modus tollens. A relatively simple maneuver, modus tollens certain does not push the human brain to its computational limits. Neither does modus tollens demand great intelligence. Nor is there any reason to believe that people fail to endorse modus tollens inferences because of a conflicting rule, a misconception. To be sure, when people hold back from endorsing a modus tollens inference, in some sense they do not understand the logic of the situation. But this is not to say they cannot understand. Nor is it to say that they hold any contrary precepts, the kinds of misconceptions highlighted above under limits of understanding.

Much the same case could be made for many departures from standard logic. While people cannot reason according to standard logic in certain senses, in other senses they can, but do not.

SHOULD PEOPLE REASON ACCORDING TO STANDARD LOGIC?

A few pages ago, I urged that the failure of the "do" clause in the Ambitious Claim raised doubts about the "can" and "should" clauses. We are adaptive organisms. If we can and should (in an adaptive sense) behave in a certain way, it's reasonable to expect that we would. But we often don't. Therefore, we either cannot or should not.

The examination of limits confirms this prediction: Indeed in several respects we cannot. However, this hardly lets the "should" clause off the hook. First of all, as just argued, important cases in which people depart from standard logic are not explained by the limits on "can". Second, recall that limits of understanding make room for learning. This also challenges the "should" clause: If people can learn what's mistaken about a precept they hold and replace it by a sounder principle of reasoning, and if the sounder principle is so much more adaptive, one would expect them to do so. Since they do not, perhaps the sounder principle is not so adaptive after all.

In keeping with this, psychologists have indeed challenged the "should" of standard logic.

Standard Logic as Contrived

One case holds that standard logic is contrived. Its formalism mismatches the kinds of practical reasoning people need to do in realistic contexts. A prime example concerns two classic errors of reasoning: denying the antecedent and affirming the consequent. As mentioned earlier, research on conditional reasoning shows that subjects commonly endorse both these implications, from 17 to 73 and from 23 to 75 percent of the time across the

range of studies reviewed by Evans *et al.* [1993, p. 36]. The question is why? Why the trend and why the variations from experiment to experiment?

One answer holds that people often interpret a conditional as a biconditional, and moreover that it often makes sense to do so. As was noted earlier, given "If the candy is blue, it's a mint," one might well suppose that the candy manufacturer color-coded the candies by flavor: blue and only blue candies would be mints. Some constructions with a conditional structure quite compellingly suggest a biconditional, for instance, "If you take out the trash weekly, I'll increase your allowance." If not, I probably won't. "If the temperature exceeds 130 degrees centigrade, the liquid will boil". If not, it probably won't.

Supporting the contextual interpretation of a conditional as a biconditional, a number of studies show a specific influence of content on the frequency of denying the antecedent and affirming the consequent. In particular, Ellis [1991], as reported in Evans, *et al.* [1993, pp. 54–55] systematically examined subjects' reasoning on eight different kinds of conditionals he defined: temporal, causal, promise, threat, tip, warning, universal, and intentional. His universal category mirrored the kinds of conditionals usually used in research on conditionals, and in this category the rates of denying the antecedent and affirming the consequent was lowest. Promises, threats, and intentions produced the highest rates of affirming the consequent and denying the antecedent. All this accords well with a practical view of people's treatment of conditionals: Why declare a promise, threat, or intention unless you do not mean to follow through should the condition fail?

By the measure of such findings, standard logic seems contrived. It would be maladaptive for people to hold back from obvious practical interpretations of conditionals for the sake of the ideals of standard logic. The consequence, indeed, would be serious misinterpretations of the contingencies in such situations. Standard logic is contrived in its pristine insistence on a context-free reading of conditional statements.

One objection to this argument observes that, although people are influenced by such practical factors, they still deny the antecedent and affirm the consequence when such factors have minimal presence—as indeed in the research cited earlier. So we cannot offhandedly dismiss the tendency toward denying the antecedent and affirming the consequent as a mere sign of practical interpretation. Another objection recalls a finding from Evans *et al.* [1993] cited earlier: The pattern of experimental results does not bear out the hypothesis that people who deny the antecedent and affirm the consequent are *consistently* interpreting the conditional as a biconditional.

These points granted, it also seems true that people appropriately supplement explicit givens from their knowledge of the world. They interpret statements by what the statements are likely to mean, not by their logical syntax. To do any less would be maladaptive.

The accusation of contrivance figures in probabilistic reasoning also. As mentioned earlier, people display a number of faulty patterns of reasoning around probability. Gigerenzer [1991] has argued that the kinds of tasks typically posed in such experiments reflect a misconceived normative model of probability and fail to tap people's true ability. Gigerenzer holds that statements about the probability of single events are meaningless. Probabilities of events are properly understood in frequentist terms like 60 times out of 100, not in single-event probabilistic terms like .6. People faced with the "Linda" problem mentioned earlier, who conclude that Linda is more probably a feminist bank teller than just a bank teller, respond to a misconceived question. They would better be asked how many out of 100 women satisfying the general description might be bank tellers, feminists, and feminist bank tellers.

There is indeed evidence that a frequentist version of the Linda question elicits much higher performance, even in the Tversky and Kahneman [1983] research. But this need not mean that the usual notion of single-event probability is misconceived. The frequentist form may simply provide subjects with a more accessible representation or may better remind subjects of what they know about probabilities. Gigerenzer's position certainly can be debated.

Nonetheless, it presents an interesting analog to the research on conditional reasoning. When one takes into account how people represent, or might best represent, probabilities, it turns out that people perform more reasonably than one might have thought. Probabilistic reasoning as typically assessed may be somewhat contrived. On the other hand, as with conditional reasoning, even taking this into account people still perform far from perfectly.

Standard Logic as Overgeneral

The contextual character of human thinking has been a prominent theme in psychological research for three decades. Considerable evidence argues that people who think especially well do so because of a large knowledge base in the domain in question. For instance, classic work on chess play conducted by de Groot [1965] and by Chase and Simon [1973] demonstrated that chess masters depend on a large repertoire of configurations of pieces that allows them to encode board positions efficiently and explore possibilities fluently. Such studies led to what has come to be called the psychology of "expertise". Subsequent work has demonstrated the importance of a large knowledge base, reflexively accessed, in physics (e.g. [Chi *et al.*, 1981]), mathematics (e.g. [Schoenfeld and Herrmann, 1982]), and other areas (in general see [Ericsson and Smith, 1991]).

This pattern of findings argues that "local knowledge" figures centrally in good thinking. General knowledge—particularly knowledge as general as

standard logic—makes only a modest contribution. The principle adaptive strategy of the human organism is to accumulate a large set of context-specific patterns through learning.

Why might this local knowledge strategy prove more adaptive than a frequent reliance on standard logic? First of all, context-specific rules express compactly what standard logic applied to more general premises could only produce through an elaborate chain of reasoning. For instance, the importance of controlling the center of the board in chess perhaps might be regenerated every time you play by reasoning according to standard logic from the rules of the game. But it's much more efficient to store such a strategic principle in memory. Second, deriving local principles through reasoning by standard logic often would strain or exceed cognitive capacity. Given the rules of chess and no experience, how easy would it be to recognize the importance of control of the center, or of configurations like pins and forks? Third, whereas the rules of chess wholly define the game, so in principle one could reason everything out, the world at large surprises us over and over again. We reason from precepts that turn out to have important exceptions. A friend turns out to be reliable only up to a point, the latest car from a reputable manufacturer proves to be a lemon, and so on. Because of the constant surprises the world serves up, nothing beats learning your way around the particulars of the situations with which you have to deal.

Such factors argue that a heavy investment in elaborate reasoning according to standard logic is a bad adaptive strategy, too hard to do and too precarious in its predictions. Better to experience, learn, and remember the ins and outs of particular situations.

This primacy of local knowledge does not stand unchallenged. For example, research on physics problem solving shows that, when faced with novel problems, successful problem solvers adopt more general strategies in tandem with their physics knowledge [Clement, 1991]. They make analogies, use their intuitions about physical situations, and deploy limiting case arguments—what would happen if this variable went to zero, infinity, or some other extreme value?

Also, people cannot simply start with the local knowledge they need. Initially, all people have to work with is whatever general knowledge they can bring to bear, including their knowledge of standard logic. In this spirit, Bereiter and Scardamalia [1993] and Bruer [1993] write of being an expert learner or an expert novice, someone who does not yet have local knowledge but who knows how to go about getting it by deploying more general knowledge.

On these and like grounds, I as well as others have argued that a strong local knowledge stance is too narrow [Perkins and Salomon, 1989], [Perkins, 1995, Chapter 9]. However, these rejoinders fall well short of restoring standard logic to the throne. The general knowledge that people demonstrably

apply is often not as general as, or otherwise not a part of, standard logic. It includes such moves as reasoning by analogy, through extreme case arguments, reasoning about promises and permissions, and so on. The patterns of reasoning have more the character of informal reasoning (discussed below) than formal reasoning. Although knowledge at many levels of generality figures importantly in good thinking, nothing recommends a heavy investment in standard logic. Good reasoning in the sense of standard logic seems to be more of a "niche skill," not the bread and butter of cognitive functioning.

Standard Logic versus Pragmatic Contexts

The theme of overgenerality takes on more specific form in a thriving line of research on what might be called pragmatic contexts of reasoning—contexts that involve promises, threats, permissions, and social reciprocity. The backdrop for most of this research is a reasoning problem called the Wason selection task [Evans *et al.*, 1993, Chapter 4], [Wason, 1966]. The Wason selection task builds upon the conditional. However, instead of asking subjects to endorse or make inferences directly, experiments ask subjects what information they would need to test a conditional rule.

For example, you are introduced to a set of cards that have letters on one side and numbers on the other. You are shown four cards and told that they might obey the rule "If a card has an A on one side, it has a 3 on the other." But do they conform to this rule? You see that the cards say A, B, 3, and 7. You need to decide which cards you should turn over to decide whether the rule is true or false.

The meaning of the conditional statement determines the correct responses. You should turn over A, because that might reveal a 7 instead of a 3, violating the rule. You should turn over 7, because that might disclose an A instead of a B, again violating the rule. There is no point in turning over B or 3, because whatever you find, the result cannot challenge the rule. However, people generally respond quite differently. They turn over A alone, or the A card and the 3 card, which cannot violate the rule. They do not turn over 7, which might challenge the rule. Correct response rates are startlingly low. In early studies employing abstract conditionals such as the one illustrated, a fully correct response occurred less than 10 percent of the time [Evans *et al.*, 1993, p. 101].

Studies manipulating various aspects of this paradigm have shown that sometimes subjects perform much better. One plausible approach holds that items like A and 3 involve too much abstraction. Perhaps people would see the logic more clearly in more concrete situations, like "If the candy is blue, then it is a mint." However, research shows that quite commonly performance proves just as poor on such versions as on more abstract versions. What does boost performance is forms of the Wason selection task that involve promises, permissions, threats, and social reciprocity.

Among many studies, an especially interesting approach shows how subjects shift their judgments according to the perspective the experimenter asks them to take. For instance, Manktelow and Over [1991] examined subjects' response to this conditional: "If you tidy your room, then you may go out to play". This announces a permission. The experiment asked subjects to view the situation from the son's perspective, checking whether mom had lived up to the rule. The subjects chose to check "tidy room" to see whether the son then got to go out to play, and "did not go out to play" to see whether the son nonetheless tidied his room. The experimenters also asked subjects to view the situation from the mother's perspective, checking whether the son lived up to the rule. In this condition, the subjects chose to check "did not tidy room", to see whether the son had gone out to play anyway, and "went out to play", to see whether the son in fact tidied the room. In other words, subjects readily handled the task framed as a matter of detecting rule violations from particular perspectives.

Gigerenzer and Hug [1992] conducted a similar study with an interesting third condition: Subjects also took the stance of an anthropologist, looking at the situation to test whether the social contract held up. Under this condition, subjects performed much more poorly.

Cheng and Holyoak [1985] sought to show that the logic of permissions, not a concrete context, played a crucial role. They tested a version of the Wason selection task featuring a conditional of the form "If one is to take action 'A' then one must fulfill precondition 'P'," and cards labeled "has taken action A," "has not taken action A," and so on. Remarkably, they found a 61 percent correct response rate compared with 19 percent on the usual abstract version. Additional work has given further support to the role of permissions in supporting performance [Griggs and Cox, 1993; Kroger *et al.*, 1993].

Several interpretations of these effects vie in the literature. One view says that people have permission schemas specialized for dealing with permissions, and likewise have schemas for other similar forms such as threats [Cheng and Holyoak, 1985]. Another view says that people have patterns of social contracts and practices of checking for cheating on such contracts built into their brains by evolution, because of the high adaptive significance of such behavior (e.g. [Cosmides, 1989]). Still another, the mental models perspective of Johnson-Laird and Byrne [1991], says that people deal with the Wason selection task and its variants by constructing mental models, much like those discussed earlier under conditional reasoning. In its abstract form, the Wason selection tasks requires multiple steps and invites mistakes; situating the task in the context of permissions or the like provokes the construction of more thorough models and supports more correct responding.

Examining which of these theories gives the best account would take us far afield—and indeed, it's not even clear to what extent the theories can be

differentiated empirically, if advocates of each are allowed to patch theirs up to accommodate new findings. Whatever the resolution of the in-fighting, studies have clearly demonstrated that a pragmatic reasoning context—a context of permissions, promises, and cheat detecting—dramatically boosts performance on the Wason selection task.

The challenge to standard logic amounts to this: People prove to be quite competent in dealing with logic in the pragmatic contexts where it counts most. Such abilities, while still quite general, point up further the overgeneral character of standard logic from the standpoint of human functionality.

Standard Logic versus Informal Reasoning

Another kind of challenge to the adaptiveness of standard logic comes from research on what is sometimes called "informal reasoning" [Voss *et al.*, 1991]. Roughly, this means contexts of inference where strict deduction and the calculus of probabilities do not easily come into play. Premises are challengeable, probabilities are hard to estimate, and so on. For instance, you may have to reason about which new car will make the best purchase in your situation, where to go on a vacation, what stocks to invest in, for whom to vote, and so on. Public issues almost always have an informal character—for example, the issue of whether recycling laws actually save resources in the end.

The simplest way to demonstrate the adaptiveness of informal reasoning would be to show that people are very good at it. No wonder, one could conclude, people are so poor at standard logic. It's not what they need to do, and they are good at what they do need to do—informal reasoning. However, research on informal reasoning disconfirms this happy conjecture. Just as people display formal errors in standard logic, they also display a wide range of poor thinking practices in informal reasoning, including the familiar *ad hominem* argument, the well-known *post hoc, propter hoc,* which means jumping to the conclusion that when one event follows upon another the first caused the second, and more. A variety of shortfalls in informal reasoning have been discussed since the days of the Greek rhetoriticians [Hamblin, 1970].

Performance aside, what can be said about the adaptive value of informal reasoning versus standard logic? One useful tack asks just how informal reasoning and standard logic relate. An answer favorable to the central position of standard logic would say that informal reasoning *is* standard logic, but in casual dress. There are premises, although they may be suppressed. If-then conditionals appear constantly, albeit in the diverse forms mentioned earlier, such as promises, threats, causal rules, and so on. Modus ponens and modus tollens come into play, although without naming or fanfare.

All this may be true. But in some respects, the structural demands of informal reasoning diverge sharply from those of standard logic. Arguments

in standard logic are characteristically one-sided, single-strand. You do not have one line of deductive argument supporting A and another not-A, given a consistent system. Moreover, if you have one sound line of deductive argument supporting A, you do not need another, except perhaps because it might be more parsimonious or elegant. Likewise, in probabilistic reasoning, you do not have one sound line of argument assigning a probability of .5 and another, based on the same set of premises, assigning .3. In fact, the rules of probabilistic inference are deductive, and any such mishap signals an erroneous inference or inconsistent premises.

In contrast, informal reasoning routinely involves two sides, with multiple lines of argument for both. Recall the issue mentioned above: Do recycling laws actually save resources? On the "pro" side, they superficially preserve resources, through recycling glass, plastic, metals, paper, and such. They may also preserve resources indirectly, by training people to be more ecology-conscious. On the "con" side, the collection and re-manufacturing processes involved in recycling themselves all consume energy and materials in various ways. Whether a net gain results becomes a complex question of ecological bookkeeping, with answers that would undoubtedly involve comparing apples and oranges. The construction of a sound informal argument thus presses reasoners to be attentive to multiple lines of inference on both sides of the case, a demand standard logic simply does not impose.

Is this distinctive structural demand of informal reasoning a problem for people? Research says yes. People show persistent "myside bias" or confirmation bias—they neglect the other side of the case, the side opposite whichever one they lean toward. For instance, a person who leaned toward the positive side of the recycling issue would likely mention two or three pro arguments and none or one con argument. In an extensive study of informal reasoning, focusing on issues like this, Perkins and colleagues found a strong tendency toward myside bias in high school, college, and graduate students [Perkins, 1985; Perkins *et al.*, 1983; Perkins *et al.*, 1991]. Baron [1991], Baron *et al.* [1993], Kuhn [1993], Means and Voss [1996], and others have reported similar findings for various populations.

One interpretation of myside bias might look to lack of knowledge on the opposite side. Another might posit some gestalt effect that impaired people from exploring the opposite side. However, further studies directly disconfirmed such hypotheses. When asked, subjects could easily elaborate arguments on the other side of the case. In fact, subjects increased the average number of arguments mentioned on the other side of the case by a remarkable 700 percent [Perkins *et al.*, 1991]. People can reason more evenhandedly; they just tend not to do so.

Not only do people neglect to search out other-side evidence, they may discount it when faced with mixed evidence. In a well-known experiment, Lord, Ross, and Lepper [**?**] explored people's positions on use of the death penalty. Before letting them see a mix of evidence on both sides of the

case, the experimenters recorded the subjects' initial positions and degrees of confidence. Then the subjects read through a deliberately mixed body of evidence that pointed in no clear direction. Ideally, the mixed evidence would reduce their confidence. To the contrary, subjects on both sides of the issue retained their positions and became even more confident. The "pro" subjects unreflectively took the "pro" evidence as reinforcing their position but tended to dismiss the "con" evidence as biased or misleading; the "con" subjects did just the opposite.

What about other kinds of shortfalls in informal reasoning? Perkins, Allen, and Hafner [1983] reported an extensive analysis of arguments from 320 people, including high school, college undergraduate, and doctoral students, as well as people who had been out of school for several years, some with just a high school and some with a college degree. Only about a quarter of the shortfalls in reasoning detected could be considered errors in formal reasoning. By and large, the shortfalls were sins of omission. Reasoners would simply neglect important considerations that might bear on the issue at hand. In other words, problems of what is sometimes called *defeasible inference* plagued people's informal reasoning: lines of reasoning could be defeated, or at least challenged, by bringing to bear further information not initially considered.

To illustrate, the most frequent shortfall was called "contrary consequent". This referred to arguments where one could start from the same premise and advance the contrary conclusion. With the recycling issue, someone might argue:

(1) If we recycle plastic, we won't have to make new plastic out of petroleum and that will preserve petroleum resources.

An example of a contrary consequent shortfall would be:

(2) If we recycle plastic, remelting the plastic and other such processes consume energy, which might well come from petroleum, depleting petroleum resources.

Of course, (2) might in turn invite further objections. The point is that it brings into the picture new information: a causal chain not considered in the original argument. Notice that (2) does not challenge any of the inferences in (1), which might well stand. Rather, it introduces another countervailing factor that brings into question the net effect of (1) and (2) on preserving petroleum resources.

Perkins *et al.* [1991] interpreted these findings in a way reminiscent of Johnson-Laird and Byrne's [1991] model theory for formal reasoning. When people reason informally, they build up a "situation model" of the target situation (which may involve alternate submodels). A situation model can include several causal chains—for instance, both (1) and (2) above—and

other relationships. The problem is, people tend to display what Perkins et al. termed a "makes-sense epistemology". People stop elaborating their situation model as soon as they arrive at one that makes superficial sense. In contrast, some people for some tasks bring to bear a "critical epistemology" that involves challenging the model-so-far, leading to revised and elaborated models that encompass more factors.

All of this amounts to an attack on the "should" clause of the Ambitious Claim. What people really should handle well is informal reasoning, the kind of reasoning life mostly requires. In particular, they should strive creatively and critically to build more encompassing situation models that reflect situations better. Handling informal reasoning well might have meant avoiding lapses of standard logic in the casual dress of informal reasoning. But research says the opposite: Informal reasoning has its own distinctive hazards. What people should do to reason better is to avoid these informal hazards, with shortfalls of standard logic a secondary priority.

THE PLACE OF STANDARD LOGIC IN REASONING

Standard logic has taken some heavy blows over the past several sections. The Ambitious Claim seems faulty on all counts. To review the case:

Do people reason according to standard logic? Very often they do not, failing to endorse even such basic deductive moves as modus tollens. Moreover, when people's reasoning performance appears to align with standard logic, they may not be carrying out logic in the usual sense. Their thinking may involve the manipulation of mental models representing possibilities rather than the exercise of inferential rules as such (although this too can be seen as rule-like, with somewhat different rules). Also, their thinking may reflect narrower rules than those prominent in standard logic, for instance rules about promises and permissions.

Can people reason according to standard logic? First of all, they cannot be expected to live up to certain ideals of standard logic, such as fully realizing potential knowledge through inference, checking consistency, or revising beliefs based on new information. Such enterprises are computationally intractable for any cognitive agent. Secondly, people's ability to wield standard logic varies considerably with intelligence, and therefore falls short of a core high-reliability performance such as managing the grammar of one's mother tongue. Thirdly, often people cannot reason according to standard logic due to misconceptions: They exercise fallacious rules and principles. They might well be able to learn better, but they have not in the natural course of events.

Should people reason according to standard logic? This too is doubtful. Firstly, in some ways standard logic seems contrived. In standard logic, statements mean only what they say; an if-then statement is not a bicondi-

tional. But in natural contexts of thinking and communication, statements mean what they most sensibly mean: People construe their likely intent and supplement them with likely scenarios, often and plausibly turning conditionals into biconditionals. Secondly, standard logic is overgeneral by the measure of people's actual adaptive reasoning strategies. Much reasoning is highly situated, a matter of knowledge suited to particular contexts. People also reason rather well in semi-general contexts, such as checking for violations of permissions and promises. Third, everyday life calls mostly for informal reasoning, which turns out to have its own traps. The classic lapses of standard logic plague informal reasoning less than myside bias and other pitfalls distinctive of informal reasoning.

Can't, Shouldn't, and Don't?

Given this barrage of evidence, one might be tempted to a conclusion directly opposite the Ambitious Claim. Far from the notion that people can, should, and do reason according to standard logic, perhaps people can't, shouldn't, and of course don't.

However, the arguments from psychology against the Ambitious Claim diminish it, not demolish it. Starting again with the "don't", it's important to recognize that some people do and sometimes most people do. As noted earlier, people deploy modus ponens reliably. Average performance on almost all standard logic tasks rises above chance level. Some people reliably score well.

Moving on to "can't", since some people usually do and many sometimes do, they can to some extent. While some people falter because of misunderstandings, people can and do learn from formal instruction, if not from just living their lives. Finally, while computational limits are real, people can still aspire to junior versions of ideals of standard logic like consistency, working to repair blatant inconsistency even as they cannot approach full-scale consistency checking.

As to the "shouldn't," no one argues that people should shun standard logic, for instance blithely affirming the consequent on a conditional that in no way invites a biconditional interpretation. The critique is weaker than that. It says simply that high investment in deliberate reasoning according to standard logic is not a good adaptive strategy. Much of the deductive reasoning we need to do occurs automatically by way of reflexive implementations of portions of standard logic [Stenning and Oaksford, 1993]. Shortfalls characteristic of informal reasoning occur more often and deserve more of our attention. In many situations, careful reasoning according to standard logic serves us poorly because of the fragility of our premises: We do better to learn about the situation directly or from others than to push our deductions to the limit. In sum, deliberate standard logic is not the best overall strategy for human reasoning moment to moment, situation to

situation. But sometimes we should, sometimes we can, and sometimes we do.

It's worth asking under what conditions standard logic speaks most pointedly to deliberate reasoning in practical contexts. And such circumstances are not hard to find. Standard logic figures centrally in mathematics, which today serves not only the hard sciences and engineering but also a plethora of fields from history to musicology. Standard logic plays an especially conspicuous role in the testing of scientific theories, not only through hypothetico-deductive reasoning but also through probability and statistical sampling theory. The branch of standard logic concerning probabilities contributes to avoiding unreasonable risks and seizing opportunities, as in gambling, making insurance decisions, performing medical diagnoses and prescribing treatments, and more.

It's true that even in such contexts people often do not reason according to standard logic. For instance, many people make insurance or gambling decisions with no knowledge of or attention to expectation values. Professional papers in psychology not uncommonly commit one or another statistical sin. But despite such "do not" phenomena, no one argues that people never do or cannot or should not. On the contrary, for their own sake and that of others, people generally should and often can and do.

On Human Rationality

Cohen [1981] saw research like that reviewed here as a misplaced attack on human rationality. How does human rationality stand today?

Let us take rationality to mean, roughly, good deliberate reasoning. Moreover, it's useful to distinguish rationality-in-prospect and rationality-in-practice. Cohen [1981] made a case for rationality-in-prospect. He held that, whatever the shortfalls people sometimes displayed, they could upon prompted reflection or with appropriate education evade those shortfalls. It seems to me that today Cohen's basic point survives rather well. While some cognitive illusions, as Cohen would like to call them, appear particularly stubborn, by and large prompting, reflection, and education demonstrably do much for reasoning (e.g. [Nisbett, 1993]).

One of the empirical challenges to Cohen's position has been the Wason selection task in its abstract form, in contrast with forms employing promises, permissions, and such. As Platt and Griggs [1993] note, most efforts to facilitate subjects' reasoning through clarifying the task and encouraging subjects to think more carefully have had limited impact. However, Platt and Griggs report an especially elaborate treatment that raised correct performance on the Wason selection task to over 80 percent. To do so, they had to offer an explication of the conditional rule, require subjects to give reasons for their judgments, and finally phrase the task as one of determining not whether the rule was true but whether it was being violated.

These tactics elevated subjects' understanding of the question and care in processing to the point where they performed largely correctly.

If empirical research does not severely challenge human rationality-in-prospect, certainly it broadens our conception of it. Considerable research on informal reasoning has shown that the prescriptions of standard logic leave much unsaid about good informal reasoning, which figures crucially in most human affairs. Rationality-in-prospect must be seen as including both good formal and good informal reasoning.

While human rationality-in-prospect may be alive and well, human rationality-in-practice continues to suffer. Twenty or thirty years ago, it was easier to say that the laboratory findings about lapses in reasoning might not mirror realistic circumstances of coping. However, considerable scholarship involving practical contexts of reasoning and simulations of them argues that in situations where it counts people often fall prey to lapses of reasoning [Baron *et al.*, 1993; Neustadt and May, 1986; Perkins, 1995; Kuhn, 1993; Tuchman, 1984; Voss *et al.*, 1991]. This does not mean that people usually reason in mistaken ways contrary to their own best interests, but that it happens often enough to worry about and to do something about through education and other means (Perkins [1992; 1995]).

Here enters one of the long-time puzzles of research on human reasoning: As well-adapted creatures, how can we get away with flawed rationality-in-practice? There seem to be two answers. First of all, as has often been pointed out (e.g. [Stich, 1990]), it's commonly the case that organisms are imperfectly adapted and nonetheless quite successful in their adaptations. This applies to reasoning as much as flying or swimming. For instance, from a Darwinian perspective the point is often not to swim or fly as fast or efficiently as possible but simply better than the competition. Evolution—and a good deal of human endeavor too—function through "satisficing" rather than optimizing, being good enough rather than perfect (Simon, [1957; 1981]). In fact, from a whole-system perspective, perfection or anything approaching it often is a poor investment of resources.

The other answer is that deliberate reasoning is only one recourse among a much wider range of adaptive strategies, including more automatic and specialized implementations of portions of standard logic and reliance on knowledge of what to do in innumerable particular areas of life where we have well-developed experience [Perkins, 1995]. Deliberate and sophisticated rationality-in-practice no doubt contributes greatly to the cutting edge of human progress, but it does not seem to be the centerpiece of day-to-day getting along.

Psychology's critique of the central place of deliberate reasoning according to standard logic has a Copernican character. Standard logic was seen as the Terra Firma of rationality, the gravitational source around which all reasoning orbited. As the Copernican revolution nudged the earth out of its central place, so the critique from psychology (and other sources too) has

pulled the ground provided by standard logic out from under our feet. What occupies the center now may be a conglomerate of principles of informal reasoning, mental models, and pragmatic contexts of reasoning. However, just as the Copernican revolution diminished but did not demolish the place of the earth, so with a revisionary view of standard logic: Less central though the place it holds in our day-to-day functioning may be, we could hardly sustain modern scholarship and civilization without it.

BIBLIOGRAPHY

[Baron, 1985] J. Baron. *Rationality and intelligence.* New York: Cambridge University Press, 1985.

[Baron, 1991] J. Baron. Beliefs about thinking. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 169-186). Hillsdale, NJ: Erlbaum, 1991.

[Baron *et al.*, 1993] J. Baron, L. Granato, M. Spranca and E. Teubal. Decision-making biases in children and early adolescents: exploratory studies. *Merrill-Palmer Quarterly*, 39, 22–46, 1993.

[Bereiter and Scardamalia, 1993] C. Bereiter and M. Scardamalia. *Surpassing ourselves: An inquiry into the nature and implications of expertise.* Chicago: Open Court, 1993.

[Braine and O'Brien, 1991] M. D. S. Braine and D. P. O'Brien. A theory of If: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182–203, 1991.

[Brody, 1992] N. Brody. *Intelligence.* New York: Academic Press, 1992.

[Bruer, 1993] J. T. Bruer. . *Schools for thought: A science of learning in the classroom.* Cambridge, MA: MIT Press, 1993.

[Byrne and Handley, 1992] R. M. J. Byrne and S. J. Handley. Reasoning Strategies. *Irish Journal of Psychology: Trinity 400 Special Issue*, 13, 111–124, 1992.

[Ceci, 1990] S. J. Ceci. *On intelligence . . . more or less: A bio-ecological treatise on intellectual development.* Englewood Cliffs, NJ: Prentice Hall, 1990.

[Chase and Simon, 1973] W. C. Chase and H. A. Simon. Perception in chess. *Cognitive Psychology*, 4, 55–81, 1973.

[Chazen, 1989] D. Chazen. *Ways of knowing: High school students' conceptions of mathematical proof.* Unpublished doctoral dissertation, Harvard Graduate School of Education, Cambridge, MA, 1989.

[Cheng and Holyoak, 1985] P. W. Cheng and K. J. Holyoak. Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391–416, 1985.

[Cherniak, 1986] C. Cherniak. *Minimal rationality.* Cambridge, MA: MIT Press, 1986.

[Chi *et al.*, 1981] M. T. H. Chi, P. Feltovich and R. Glaser. . Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121–152, 1981.

[Clement, 1991] J. Clement. Nonformal reasoning in experts and in science students: The use of analogies, extreme case and physical intuition. In J. Voss, D.N. Perkins, & J. Segal (Eds.), *Informal Reasoning and Education*, 345–362. Hillsdale, NJ: Lawrence Erlbaum Associates, 1991. [Flexible expertise]

[Cohen, 1981] L. J. Cohen. Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences*, 4, 317–370, 1981.

[Cosmides, 1989] L. Cosmides. The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31, 187–276, 1989.

[de Groot, 1965] A. D. de Groot. *Thought and choice in chess.* The Hague: Mouton, 1965.

[Ellis, 1991] M. C. Ellis. *Linguistic and semantic factors in conditional reasoning.* Unpublished PhD thesis, University of Plymouth, 1991.

[Ericsson and Smith, 1991] K. A. Ericsson and J. Smith, eds. *Toward a general theory of expertise: Prospects and limits.* Cambridge: Cambridge University Press, 1991.

[Evans *et al.*, 1993] J. St.B. T. Evans, S. E. Newstead and R. M. J. Byrne. *Human Reasoning: the psychology of deduction.* Hove, UK: Lawrence Erlbaum Associates, 1993.

[Flamagne, 1975] R. J. Falmagne, ed. *Reasoning: Representation and process in children and adults.* Hillsdale, New Jersey: Erlbaum, 1975.

[Gardner, 1983] H. Gardner. *Frames of mind.* New York: Basic Books, 1983.

[Gigerenzer, 1991] G. Gigerenzer. How to make cognitive illusions disappear: Beyond "heuristics and biases." In W. Stroebe, & M. Hewstone (Eds.), *European review of social psychology, Volume 2.* New York: John Wiley & Sons, 1991.

[Gigerenzer and Hug, 1992] G. Gigerenzer and K. Hug. Domain-specific reasoning: Social contracts, cheating and perspective change. *Cognition*, 43, 127–171, 1992.

[Gould, 1981] S. J. Gould. *The mismeasure of man.* New York: W. W. Norton, 1981.

[Grice, 1975] P. Grice. Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Studies in syntax, Vol 3: Speech acts.* New York: Academic Press, 1975.

[Griggs and Cox, 1993] R. A. Griggs and J. R. Cox. Permission schemas and the selection task. In J. St.B. T. Evans (Ed.), *The cognitive psychology of reasoning* (pp. 637-651). Hillsdale, NJ: Erlbaum, 1993.

[Hamblin, 1970] C. L. Hamblin. *Fallacies.* London: Methuen, 1970.

[Harman, 1986] G. Harman. *Change in view: Principles of reasoning.* Cambridge, MA: MIT Press, 1986.

[Henle, 1978] M. Henle. Foreword. In M. Revlin & R. E. Mayer (Eds.). *Human reasoning.* Washington, DC: Winston, 1978.

[Herrnstein and Murray, 1994] R. J. Herrnstein and C. Murray. *The bell curve: Intelligence and class structure in American life.* New York: The Free Press, 1994.

[Jensen, 1980] A. R. Jensen. *Bias in mental testing.* New York: The Free Press, 1980.

[Johnson-Laird, 1983] P. N. Johnson-Laird. *Mental models.* Cambridge, Massachusetts: Harvard University Press, 1983.

[Johnson-Laird and Byrne, 1991] P. N. Johnson-Laird and R. M. J. Byrne. *Deduction.* Hove, UK: Lawrence Erlbaum Associates Limited, 1991.

[Johnson-Laird *et al.*, 1992] P. N. Johnson-Laird, R. M. J. Byrne and W. Schaeken. Propositional reasoning by model. *Psychological Review*, 99, 418–439, 1992.

[Kahneman and Tversky, 1982] D. Kahneman and A. Tversky. On the psychology of prediction. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 48–68). Cambridge, UK: Cambridge University Press, 1982.

[Kahneman *et al.*, 1982] D. Kahneman, P. Slovic and A. Tversky. *Judgment under uncertainty: Heuristics and biases.* Cambridge, UK: Cambridge University Press, 1982.

[Kroger *et al.*, 1993] J. K. Kroger, P. W. Cheng and K. J. Holyoak. Evoking the permission schema: The impact of explicit negation and a violation-checking context. In J. St.B. T. Evans (Ed.), *The cognitive psychology of reasoning*, pp. 615–635. Hillsdale, NJ: Erlbaum, 1993.

[Kuhn, 1993] D. Kuhn. Connecting scientific and informal reasoning. *Merill-Palmer Quarterly*, 39, 74–103, 1993.

[Lindman and Edwards, 1961] H. G. Lindman and W. Edwards. Supplementary report: Learning the gambler's fallacy. *Journal of Experimental Psychology*, 62, 630, 1961.

[Manktelow and Over, 1991] K. I. Manktelow and D. E. Over. . Social roles and utilities in reasoning with deontic conditionals. *Cognition*, 43, 127–171, 1991.

[Means and Voss, 1996] M. L. Means and J. F. Voss. Who reasons well? Two studies of informal reasoning among children of different grade, ability and knowledge levels. *Cognition & Instruction*, 14, 139–178, 1996.

[Miller, 1956] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63, 81–87, 1956.

[Neustadt and May, 1986] R. E. Neustadt and E. R. May. *Thinking in time: The uses of history for decision-makers.* New York: Free Press, 1986.

[Nisbett and Ross, 1980] R. Nisbett and L. Ross. *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, New Jersey: Prentice Hall, 1980.

[Nisbett, 1993] R. E. Nisbett, ed. *Rules for reasoning.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.

[Nisbett *et al.*, 1993] R. E. Nisbett, D. H. Krantz, C. Jepson and Z. Kunda. The use of statistical heuristics in everyday inductive reasoning. In R. Nisbett (Ed.), *Rules for reasoning* (pp. 15–54). Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.

[Perkins, 1985] Perkins, D. N. (1985). Postprimary education has little impact on informal reasoning. *Journal of Educational Psychology*, 77, 562–571, 1985.

[Perkins, 1992] D. N. Perkins. *Smart schools: From training memories to educating minds.* New York: The Free Press, 1992.

[Perkins, 1995] D. N. Perkins. *Outsmarting IQ: The emerging science of learnable intelligence.* New York: The Free Press, 1995.

[Perkins and Salomon, 1989] D. N. Perkins and G. Salomon. Are cognitive skills context bound? *Educational Researcher*, 18, 16–25, 1989.

[Perkins *et al.*, 1983] D. N. Perkins, R. Allen and J. Hafner. Difficulties in everyday reasoning. In W. Maxwell (Ed.), *Thinking: The frontier expands* (pp. 177189). Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1983.

[Perkins *et al.*, 1991] D. N. Perkins, M. Farady and B. Bushey. Everyday reasoning and the roots of intelligence. In J. Voss, D. N. Perkins, and J. Segal (Eds.), *Informal reasoning* (pp. 83–105). Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1991.

[Platt and Griggs, 1993] R. D. Platt and R. A. Griggs. Facilitation in the abstract selection task: The effects of attentional and instructional factors. In J. St.B. T. Evans (Ed.), *The cognitive psychology of reasoning* (pp. 591–613). Hillsdale, NJ: Erlbaum, 1993.

[Revlin and Mayer, 1978] R. Revlin and R. E. Mayer, eds. *Human reasoning.* Washington, DC: Winston, 1978..

[Rips, 1990] L. J. Rips. Reasoning. *Annual Review of Psychology*, 41, 85–116, 1990.

[Rips, 1994] L. J. Rips. *The psychology of proof: Deductive reasoning in human thinking.* Cambridge, MA: MIT Press, 1994.

[Roberts, 1993] M. J. Roberts. Human reasoning: Deduction rules or mental models, or both? In J. St.B. T. Evans (Ed.), *The cognitive psychology of reasoning* (pp. 570–589). Hillsdale, NJ: Erlbaum, 1993.

[Ross and Anderson, 1982] L. Ross and C. A. Anderson. Shortcomings in the attribution process: On the origins and maintenance of erroneous social assessments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 129–152) Cambridge, UK: Cambridge University Press, 1982.

[Schoenfeld and Herrmann, 1982] A. H. Schoenfeld and D. J. Herrmann. Problem perception and knowledge structure in expert and novice mathematical problem solvers. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8, 484–494, 1982.

[Simon, 1957] H. A. Simon. *Models of man: Social and rational.* New York: Wiley, 1957.

[Simon, 1981] H. A. Simon. *The sciences of the artificial.* Cambridge, Massachusetts: The MIT Press, 1981.

[Smith *et al.*, 1993] E. E. Smith, C. Langston and R. E. Nisbett. The case for rules in reasoning. In R. Nisbett (Ed.), *Rules for reasoning* (pp. 361–397). Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.

[Stenning and Oaksford, 1993] K. Stenning and M. Oaksford. Rational reasoning and human implementations of logic. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 136–176). London: Routledge, 1993.

[Stich, 1990] S. Stich. *The fragmentation of reason: Preface to a pragmatic theory of cognitive evaluation.* Cambridge, Massachusetts: MIT Press, 1990.

[Tuchman, 1984] B. Tuchman. *The March of Folly.* New York: Alfred A. Knopf, 1984.

[Tversky and Kahneman, 1982] A. Tversky and D. Kahneman. Judgments of and by representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 84–98) Cambridge, UK: Cambridge University Press, 1982.

[Tversky and Kahneman, 1983] A. Tversky and D. Kahneman. Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological review*, 90, 293–315, 1983.

[Voss *et al.*, 1991] J. F. Voss, D. N. Perkins and J. W. Segal. *Informal reasoning and education.* Hillsdale, NJ: Erlbaum, 1991.

[Wason, 1966] P. C. Wason. Reasoning. In B. M. Foss (Ed.), *New horizons in psychology,* Vol. I. Harmondsworth: Penguin, 1966.

ELSE M. BARTH

A FRAMEWORK FOR INTERSUBJECTIVE ACCOUNTABILITY: DIALOGICAL LOGIC

Although dialogical logic — or, dialogue logic — offers a radical change in most concepts of logic, papers have usually dealt mainly with its abstract features. The present exposition wants to emphasize its enormous potential for practical logic. For this purpose it is necessary to distinguish dialogical logic both from the group of post-medieval traditional logics and from modern (roughly, twentieth-century) logics. Dialogical logic offers a complete re-modelling of the latter, whereas traditional logics are not amenable to remodelling in dialogical terms. This paper brings a theoretical survey of dialogical/dialogue logic and reviews its categories in terms of their traditional predecessors, delving into the history of ideas for a deeper assessment of the importance of this framework. A study of the history of ideas concerning logic provides insight in the need for logical instruments that take human dialogue openly into account.

Each section opens with a survey of the titles of sub-sections, which may be used as an extension of this summary.

I. Critical dialogue – on a backdrop of "Being"?

1 ONTOLOGY AND ACCOUNTABILITY

A plea for applied logic? — Critical discussion — Anti-parliamentary philosophy and the logic of impetus — Accountability — Semiotics according to strong elitism — Being — Knowledge and the bearers of Logos — Confidence, certainty and intuition — Criticism and reflection of the General — Two-ity and opposition — Negation

1.1 A plea for "applied logic"?

Practical logic is the topic of this book and the nature of dialogical logic the topic of this survey. We may therefore disregard metalogical studies of a mathematical kind, as well as the professional interests of mathematicians. Only when mathematical ideas touch — as they often do — on practical notions and interests shall we mention them.[1]

Nevertheless, there is as yet no point in going directly for applied logic. The concept is not yet really meaningful: which logic? The science of logic first has to become applicable, which in more senses than one it still is not. Under the present

[1] Moore and Hobbes [1996] may be mentioned here.

circumstances — as discussed in Albert's sophisticated 1996 paper on frameworks — logic still has to be *made* applicable to international dialogue.

A few years ago the slogan: "Keeping the discussion going" was launched as a philosophical imperative (Rorty). It is a fine slogan. However, saying it is not enough and should not be the last stage of philosophy of any kind. In order to enhance the chances that, little by little, discussions will become progressively more frequent, effective and global, instruments for discussion must be improved, extended, developed.

1.2 Critical discussion

Critical discussion is a subcategory of dialogue. In any century a great number of persons have eagerly pursued "rational" discussion with other minds. Though they must have felt that not all arguments are acceptable, in earlier centuries there never was any clear and systematical link between "rational discussion" and "logic". This meant that even in our own century logic — whether understood as deductive, constructive, two-valued, or whatever — could be and often was (mis)understood as a serious stumbling-block to the dialogical project. Conversely, this seemed to justify a continued view on "Logic" — the one and only — as representing "higher" ambitions and objectives. Logic was, *ipso facto*, unsuited to practical use.

Here we have localized one of the several motives behind anti-parliamentary philosophies.

The spirit of dialogue logic may seem closer to that of justice and legal rights than to theology or mathematics. However, in dialogue logic no accusation of guilt is involved, and it may therefore be better still to take political and scientific parliamentarism as sources of orientation.[2]

1.3 Anti-parliamentary philosophy, and the logic of "impetus"

Between parliamentarism and the Logic of Being there is an inverse relation. "Logic is for us to questioningly review the Grounds of Being, the very abode of questionableness," said Martin Heidegger, a member of Hitler's NSDAP party since 1933, in 1934. Heidegger definitely does not connect logic with criticism by political opponents. Academic freedom, he said in 1933, must be cast off, it is "only negative" [Kiesewetter, 1996, p. 268]. The same inverse relation is found between parliamentarism and the philosophy of Imperial Minds in Intuitionism, a philosophy of mathematics deriving from German idealism and sporting an anti-dialogical epistemology of "the Creative Subject" (Brouwer).

[2]Jean-Paul van Bendegem and Jan Albert van Laar have kindly read and commented on versions of the manuscript, and so has Ralph Johnson. The responsibility for omissions and shortcomings is my own. Henk Misset as well as my sometime co-authors Erik Krabbe, Jo Martens and Rob Wiche also all deserve my personal acknowledgement in the present paper. I have chosen to use roman type for pragmatical variables throughout.

Sixty years ago the appeal to the intuitions of the Creative Subject was still common in literature, mathematics and practical politics — a return to the medieval logic of impetus (a term in medieval dynamics, standing for a hypothetical force supposedly inherent in any moving body or development, and eliminated from mechanics by Galileo (who was also a great debater) and Newton [Finocchiaro, 1980].)

In matters of logic some authors see a dialogical development in the Middle Ages, disrupted in the thirteenth century by the Goddess of Reason and the logic of Peter of Spain. In Peter's concept of Reason the key term is *fides*, meaning *confidence*, *trust* [Ong, 1983, p. 65]. Other authors — Bird, Jaspers, Moody — mention still earlier sources: Augustine and gnostic-neoplatonist patristics, Avicenna, Boethius.

New, more practical forms of logical thought and behaviour have since developed in science as well as in parliamentary practice. Outside these fields of activity the newer forms are not much used. This is in part due to a devotion to "pure mathematics" in mathematical-logical circles, in part to fundamentalist assumptions in religion, theology, and politics. Logical theory was often systematized by thinkers with a background in theology. The incomparable influence, even in this century, of St. Augustine as a source of anti-dialogical reason mostly goes unrecognized. An invitation to seek Understanding and Truth through introspection is found in all divisions of theology, if not in all texts, and often in general philosophy as well. It combines with an ego-centred semiotics as displayed in, e.g., the works of Leibniz (see the study by Prior). In the eighteenth century a broader spectrum of logicians became involved, but in the twentieth the systematization of logic became once more dominated by pure (rather than applied) mathematicians demonstrating aspirations verging on the theological, often with roots in gnostic-type, interiorist philosophies of mathematics.

However, the dialogical project in general philosophy is old, say some (referring to Plato's dialogues). It is new, say others. Its rebirth is sometimes set to the beginning of the century, with a first culmination point in [Buber, 1923]. Buber, however, was not a logician.

Why is it so important to adopt the dialogical approach in logic? In what kind of connections is this particularly pressing?

Once a dialogical outlook is adopted, an abundance of theories and weird entities that once were called upon by logicians become demonstrably superfluous at best, if not downright damaging to the project. Examples range from mathematics to homeopathy and from linguistics to political thought.

1.4 Accountability

The dialogical approach, in anything called logic, means that the idea of accountability in intersubjective affairs is understood as the primary goal. This implies

that debaters have obligations.[3]

Resistance to a new outlook will always be strong, and no less so where philosophical dogma combines with positions of political power. A highly influential part of our semiotic heritage is linked to the bloody dramas of our own century. It is at war with that of critical dialogue:

1.5 Semiotics according to strong elitism

Strongly elitist systems of semiotical assumptions are those which distinguish operationally between an elite and the masses or the people. This incorporates semiotical dogmas of fascists, Nazis, revolutionary conservatives and communists. Members of the elite are those who co-operate with *the development* of history/society, helping it forward by helping God create. Barring the highest members of the elite, individuals are of little interest, the important units being larger collection (totalities) and their spokesmen. A disagreement within the unit is important *only* as a conflict of *motives, interests or values*. It is not scrutinized as a conflict of avowed opinions. From pre-WW II elitism dates the expression "negativism and criticism". A critical attitude is a socially negative phenomenon, being *non-creative* as well as lacking in solidarity: critical minds *enfeeble the elite* in its creative activities. The political focus is on *decision* and on *action*; the members of the elite are capable of *creative decisions*. This older concept of decision is not taken from the theory of choice but from the philosophy of elitist spirituality.

Suggestion

Confronting the masses (political subjects, the non-initiated), the elite employs suggestion. Those who cannot belong or do not yet belong to the elite should *give themselves up to* its representatives and the prevailing order. One individual can *represent* the elite as well as the interest of the masses.

"Being"

As the authors who recommend this strong elitism know it, philosophy is concerned with the study of Being (ontology). Every norm, law, or theory, be it religious, physical or logical, has a firm ground (fundament) in Being. The ground itself is "something", i.e., Being has at least one *substance* — spirit, matter, or energy — together with *essences* suspended in the substance. Into this ontology the elite-mass ideology translates as follows: "Being" harbours *ontological duality* which is *reflected* in a social duality. Extreme modes receive particular attention. The frequency of "two" is high; in Europe as well as in China this is connected with *Yang and Yin*. Statements about ontological polarities abound, with particular emphasis on *enemy-friend*.

[3]Or, should imply. Lorenz and Barth emphasize obligations in addition to rights, whereas Krabbe has pleaded for the conceptual sufficiency of the concept "rights".

The category pair *statemental proponent-opponent* is either absent from the discourse or else fallaciously reduced to that of *enemy-friend*. When history and society are amenable to (divine) development, Divine Being is seen as the *goal* — also of "Logic". "Nothing", or "das Nichts", which is mixed into it, is an ontological, not a dialogical notion. Any truly new step is, so the story goes, created *ex nihilo*.

1.6 "Knowledge" and the "bearers of Logos"

In every problem sphere one assumed the existence of *fundamental laws* (such as the *Law of the Ground* itself). The light of the spirit allows the elite *direct insight* into essences. By direct insight they discern what is *general* in every individual of a given *species*. This esoteric knowledge is aspired to in the form fundamentally self-evident to some, yet inter-subjectively valid insights. A member of the *elite* has been given the gift of the spirit, is the *bearer of Logos/collective thought* and therefore *represents the Logical/collective consciousness*. What is probably the most systematic attack on this evidentialism comes from a logician [Beth, 1959a].

1.7 "Confidence", "certainty", and "intuition"

Until the twentieth century the set of basic epistemological categories did not normally include criticism. *Verbal* criticism and opposition was morally bad, a sign either of destruction or of vacillation. Confidence (*fides*) and Certainty were expected and much appreciated attitudes, notably the certainty of "the Genius". In the elite, Knowledge is Wisdom. Trust is invested in what is old, classical. In subjectivist versions of elitism the *pure I* (in the singular) may reach truth by "intuition", which is a matter of insight into the *interior* of the genius. The genius intuits both "two-nesses" and "two-oneness" (both to be understood ontologically, and not semiotically).

Among the components of Fascist and Nazi mind-sets is the emphasis on confidence, trust, belief, and co-operation, both with authorities and with "the Tendency of the World". Critical debate was held in deep contempt.

1.8 Criticism and "reflection of the General"

The important verbs, in elitist epistemology, are to *intuit*, to *judge*, also called to *reason*; in Logic, to *be*, to *reflect* (*mirror*). The science of logic then is a *Logic-of-being*, or a *Logic-of-the-spirit* concerning *the pure I*, or both at the same time. The elite's *reflection* results in an image, i.e., it equals that which it reflects.

Members of the elite can discern "what is general — the essence, the type — *in* an individual object, event, or case. There is an assumption of a *general nucleus* in each and every "phenomenon". The nucleus has been called by some *an indefinite, floating individual* (Maritain), by others *a logos-bearer* ("*logophore*", von Freytag). An indefinite, floating nucleus and bearer of the Logos in question

is taken to *represent* a “species” and all its members. The elite directly and surely discerns generalities ranging from *the Triangle* — Augustine’s example — to *the Bantu*, *the Jew*. For the genius and for the elite, the reflected image entails direct, immediate, “intuitive” conviction.

No critical discussion is needed, we are confronted with *a pre-judicium*. The only further step is to (re-)present the image to *a passive audience*, by suitable means: anecdote, legend, or myth. In this process only the intuiting logos-bound mind, now in the role of reporter, is at stake, and no one else. Members of the elite may themselves be understood as personifications of *the logophore in the People* (this is a part of the logic of fascism, too).

Does this indicate that long before Watson and Crick the intellectual tradition had already grasped the importance of DNA and inherited traits? Triangles are not in the possession of DNA, hence this cannot be the explanation.

1.9 “Two-ity” and “opposition”

The concept *two* is, however, dominant as well. One attempts to reduce all non-hierarchical relations to *equality and difference*, both binary and symmetrical. *Hierarchical order* is the one and only acknowledged kind of asymmetry; this makes for an enormous emphasis on hierarchical notion, also in semiotics. *Dynamics* is provided by an ontological-conceptual category called “opposition” or ”contradiction”.[4] This category, which is usually understood antagonistically, was fundamental and remained undefined. *Difference, fraction, opposition, relation, relativity*, even *conflict* are still taken as synonyms by many who write on philosophy or on practical matters. A distinction between antagonistic and non-antagonistic *opposition* is sometimes found, but not often. The ontological theory of *change* and *development*, called *dialectics*, starts from the assumption of two antagonistic somethings and involves a stride, a gradual or ultimate *union of the two*, and a new *birth*.

1.10 Negation

A slow change can be discerned. Many publications connect negation with dialogue, as perhaps the most dialogue-based of all logical categories (cf. [Apostel *et al.*, 1972; Apostel, 1972; Hoof and Hoepelman, 1991; Wiche, 1993; Hoof, 1995]. This was not always so. Ontology-based logic features “negation of Being(s)”; negation in thought or in language is then usually treated as binary, resulting from an opposition in Being, hence close to ideas of annihilation.

This makes for a weak use of “not” in verbal intercourse; in full harmony with the morals of *co-operation* with the development of Being (the Universe, history, or society). Critical discussion is then held in contempt, the focus being on non-verbal decision. As to speech acts the only good ones are *to affirm, to trust*, and to

[4] And sometimes “contrariety”, as in [Cornford, 1957, 65f, 69ff].

keep silent. To *deny* (negate) a statement counts as nothing else than to not affirm. [5]

In such a semiotical surrounding the inference theory concerning unary negative particles, e.g. “is false”, “is not the case”, becomes extremely weak. This does not matter as long as one trusts the elite’s intuition of what is general and essential, as accompanied by anecdote and myth as means of suggestion.

The role of opponent-in-debate is not reflected in this mind-set; attention to the *interplay in debate* of speech acts is non-existent.

One of the first authors to systematically take up this fact was Hans Lenk, in his magisterial [1968] — a rich book combining logic and the history of ideas but, unbelievably, never translated into English (see below). I would like to think that the spirit of that work is to some extent reflected in the present paper as well as in earlier publications of mine.

2 A NON-DEDUCTIVE FRAMEWORK

The adaptation problem in logical theory — A non-deductive framework — Modern logic — Sources — Terminology — Pragmatism

2.1 The adaptation problem in logical theory

The philosophies of Reason therefore, whether individual or absolute, almost entirely exclude intersubjective *dis*-agreement and intersubjective *mis*-understanding from consideration. Understanding and Agreement, hence also misunderstanding and disagreement, were understood non-dialogically, either as a relation of one mind to a phenomenon or as concerning sheer introspection. When legend and myth are idolized as they were in the first half of this century, clarity and interpretability become negative features and do not deserve intelligent inquiry.

Dialogical logic takes care of disagreement among human individuals by the very concept of logic it represents: reasoners and arguers are offered concepts, advice and possible norms that derive not from Being but rather from human experience and human institutions. The same holds of another category that was disregarded, to our peril, by most theorists, including David Hume: human *mis*-understanding.

A practice-oriented logic, conceptually attuned to the diversity of human minds and able to take care of these differences by verbal means, is needed for purposes of many kinds — from politics to the physical sciences. Could not such a logic be brought about? Even in this century “logic” has — by definition — been understood as “deductive”. It was also “formal”, later improved to “formalized”. “Deductive” and “formal” were taken not exactly as synonyms, but no one gave

[5]To my knowledge, in modern times Naess was by far the first to introduce speech acts into an informal dialogical approach to logic, as Lorenzen was the first to connect speech acts with formal approaches to logic. Naess did so in lectures and publications since 1940, Lorenzen in a lecture at an international congress in 1958, published in 1960.

much explanation of their meanings either (we shall return to concepts of "formality" below). As for science, the demands of intersubjective understanding and of critical discussion among scientists have always formed part and parcel of their business, if not always of their definition of "logic". In politics, parliamentarism has been known and applied for some time in many parts of the Western world. Here we shall be concerned with the problem of *adapting* the "deductive" science of logic generally — whether "formal" or "informal" — *to the practical world*.

2.2 A non-deductive framework

Dialogical Logic, or DL, for short, is a non-deductive theoretical framework for the further development and refinement of human logic as a *logic of social accountability in practice*.

2.3 "Modern logic"

Notice that elementary modern logic, descried as a "deductive" science of a "deductive" phenomenon, has already been brought into dialogical form. This is a fact of great importance. However, DL should not be identified with the result (call it DL*) of that transformation; rather, DL is a general framework for human logic in the wider sense, as an open pursuit. The relation of "deductive logic" to dialogue logic is therefore not that of the former to, say, modal logic. As an extension of elementary deductive logic, modal logic offers rules for deduction featuring certain logical operators, "necessary" and "possible", that elementary logic ignores. In fact modal logic can be rephrased as a proper part of dialogical logic (briefly discussed below).

Though DL certainly does introduce new argumentational categories, it does not start from new logical particles. In other words, DL is not basically a set of new rules for deduction in the neglected field of "discussion". It is an alternative framework for logic in general that allows us to dispose of old impractical conceptions so as to make logic ready for practical life.

Also, it is not ready-made for all times. Though it does offer a great many practical suggestions and possibilities that cannot be had elsewhere, it is not yet a panacea for every problem that asks for an application of a logic. A fortiori this holds of the dialogical transformation DL* of "modern" logic. What, then, is its value? Its power entirely to replace without loss the conceptual world of "deduction" will not be fully demonstrated here; see the papers in [Lorenzen and Lorenz, 1978; Barth and Krabbe, 1978; Barth and Krabbe, 1982; Lorenz, 1992], with references to publications by other authors. Here the following remarks are relevant.

DL presents an approach to "logic" and "reasoning" which, right at the starting-point, sets reasoners and arguers down at a considerable distance from the solipsist atmosphere of the isolated Ego (or computer), as well as from the objectivist atmosphere of an individual "I" confronting "the world" — two images that together

have characterized "Reason" in the Western hemisphere since time immemorial (cf. [Prior, 1977]).

2.4 Sources

The basic aspects of dialogical logic presented below are primarily indebted to the philosopher-logicians Arne Naess, Evert Willem Beth,[6] and Paul Lorenzen.

The further pragmatical and operational embedding is indebted also to P. W. Bridgman, Pierre Duhem, Pierre Poincaré, Arend Heyting, Richard Montague, and Walter J. Ong. There are affinities between the dialogical outlook on logic and the debate-oriented philosophies of C. L. Hamblin and Nicholas Rescher; of Karl Popper, Hans Albert and Ernst Topitsch; of Ludwig Wittgenstein and J. L. Austin, and, let us hope, Richard Rorty.[7]

2.5 Terminology

The technical vocabulary below is a development of that of Lorenzen (originating in [1960]), Lorenzen and Lorenz (see also [1978]), Barth and Krabbe, to which is added some of Naess' notation.

Terms for dialogical roles and other pre-eminent categories have been given capital initials, as an indication of their status as basic terms (and in our descriptions of anti-dialogical philosophies we have done the same with the basic terms thereof).

2.6 Pragmatism

In connection with general logic it is not only of importance to distinguish between "two-valued" and "constructivist" approaches. For an understanding of dialogue-logic and its potential it is at least as important to grasp the difference between creationist philosophies and pragmatist-constructivist philosophies, most importantly: between *Creationist-intuitionist constructivism*, which is explicitly anti-dialogical, and *Dialogical-pragmatist constructivism* (Lorenzen's Erlangen school a.o.). This terminology — which is not meant to reflect a dichotomy — will be used below.[8]

---

[6] See my [1990] and [Barth *et al.*, 1999].

[7] Hintikka's game-theoretical theory of *semantics*, formally$_2$ isomorphic both with Beth's and Lorenzen's approaches to logic from the nineteen-fifties and sixties, is represented by his own chapter in the present volume.

[8] See the definition and discussion of "radical constructivism" in[Glasersfeld, 1998] and the ensuing discussion in the *EuS* (cf. the contributions by Janich, Taschner, Weber, Zahn).

II. Can we construct an informal dialogue logic?

3 TRANSFORMATION OF EXTANT THEORIES

Three kinds of formality, and the rejection of forms$_1$, with chart — What is the problem situation? — Solving the adaptation problem (1). To pragmatize, acts 1–4 — To pragmatize, Def. 1–3 — States of activity: Def. 4 — Weak and strong pragmatization, Def. 5 — Sources of logical (in)validity: problem-solving capacity and institutional conventions, act 5, Def. 6

3.1 Three kinds of "formality"

Misunderstandings about "forms" abound. To an exposition of DL it is crucial to distinguish between the widely different meanings of "form" In Western conceptions of logic the words "form" and "formal" have at least three, quite different, meanings.

Under *form*$_1$ we shall subsume Platonic *ideas*, or *formae*, Aristotelian *essences*, as well as ontological *substances*. Hence by *formal*$_1$ we shall understand what pertains to such *formae*, essences, or substances. (Observe that the word "formalization" is never used in this connection.)

An *informal*$_1$ contribution to our understanding of logic is one that that does not presuppose such forms or essences.

By a *form*$_2$ we shall understand a syntactic norm in an (internal or external) language, provided the norm is defined by the use of non-referential categories used in discussion, symbolized by so-called "logical particles". By *formal*$_2$ we shall understand what pertains to such syntactic forms$_2$. *Formalization*$_2$ is then the step of making the forms$_2$ quite explicit. An *informal*$_2$ contribution is one that does not aim at doing so.

By a *form*$_3$ we shall understand any rule other than the syntactic ones of language or use of language and cognition, and by *formal*$_3$ what pertains to certain forms$_3$. Informality in this sense is hard to come about today, though some may still speak of "logic", as in the old tradition of situating "logic" within what they call "Being".

Then there are, in principle, also three kinds of "formality", hence three kinds of "formalism". In fact,

- *pre-modern logic* was almost always formal$_1$, extremely informal$_2$, its formalities$_3$ being remarkably meagre. Most of the philosophical logic that has been applied in practice is of this nature;

- *Fregean and post-Fregean theoretical logic* ("modern logic"), on the other hand, is entirely informal$_1$ (though among its theoreticians and practitioners quite a few give the impression that they would have liked this to be different). It is mostly put in formal$_2$ terms and only in such terms. This opens for a strengthening of the formal$_3$ apparatus.

It would be a serious mistake to think that formal$_3$ rules require advance formalization$_2$. However, to many thinkers the ambiguousness of "form" has been confusing and misleading, suggesting to them that all ideas, practical advices and practical morals of Fregean or post-Fregean logic require a high degree of formalization$_2$.

Proviso: In the title of this chapter the word "informal" stands for "informal$_1$ as well as informal$_2$", or, more briefly, "informal$_{1,2}$".

The rejection of forms$_1$ took place in "deductive" logic in the late 19th-early 20th century. The three widely different meanings of "form" show that in distinguishing "formal" from "informal" we must consciously operate with 3 dimensions. Here an overview may be convenient:

"Deductive" logics	formal$_1$	formal$_2$	formal$_3$
Aristotelian	+	+ weakly	+
Most pre-1900, pre-dialogial	+	-	-
Post-1900/Fregean, pre-dialogical	-	+	+

So "deductive logic" has vastly different meanings, too.

The dialogical version of modern "formal" logic — a formal$_2$ DL — may be called DL*. All that first-order formal$_2$ logic can do for you, DL* can do, too, and more.

We shall see that part of the framework to be used for this purpose has a systematical as well as suggestive value far beyond DL. We assume that a strong pragmatization of deductive logic is needed.

Speaking generally, the construction of DL will be in terms of formal$_3$ but informal$_1$ rules and definitions. In our reconstruction of DL* we shall have to add formal$_2$ rules and definitions.

3.2 *"What is the problem situation?"*

A purpose of any pragmatization is *to disclose the problem situation* which the theory is to serve. This requires close knowledge of the theory in question.

For the purpose of adapting logic to the practical world, a *pragmatization of logic itself* — the concept, the rules, and the way of bringing them together in a systematic manner — is clearly required. This process is far from completed, though a successful pragmatization of formal$_2$ modern logic has indeed been carried out.

3.3 *Solving the adaptation problem (1).*

"To *pragmatize*". Notice that we do not define an adjective "pragmatical" nor a noun "pragmatism", these being terms which belong with classificatory tasks and

with such tasks only. We want to achieve a change rather than offer a classification, hence we need a verb, and this verb should be well defined. We may use the verb "*to pragmatize*". From this word we can construct a process name: "pragmatization", which is to stand for the systematic, stepwise *transformation of extant theories, scientific or philosophical,* so as to bring to the fore one or more hidden "practical" coordinates and variables, which often are found stowed away as "context". (A more precise definition is Definition 2 below.)

(Act 1) Whereas objectivist logic emphasizes several kinds of *referential meaning* of terms (such as reference to the outer world, or to concepts), all pragmatical approaches will emphasize their *meaning-in-use*, but they may do it differently.

(Act 2) One element in a process of pragmatization, of whichever theory, is to systematically insert *operators' coordinates* or *operators' roles* and to re-write definitions, hypotheses, rules and results accordingly. This may or may not lead to a more powerful theory. Einstein's pragmatization of Newtonian mechanics is a more powerful theory than Newton's for celestial velocities.

In logic this means that we leave behind objectivist no-role as well as subjectivist one-role philosophies in favour of a theory of *meaning-in-use*, involving *logical roles*.

(Act 3) Another component of the pragmatization e.g. of logic is the step from "deduction from first principles" to a *cost-benefit assessment of logical theory itself*. (The notion of *problem-solving* capacity should be defined partly in terms of cost-benefit analysis, also when logic itself is concerned.)

(Act 4) Yet another component is the step from pure rationalism over to empiricism in the philosophy of logic(s).

The noun "pragmatization" may be useful as a name for the process as a whole.

DEFINITION 1. To *pragmatize* a given theory means to eliminate non-practical terms relating to *"Being"* through systematical insertion of semantical and logical *roles* and re-writing definitions, hypotheses, rules, methods and results accordingly.

DEFINITION 2. The noun "pragmatization" will be used for any process that encompasses Acts 1–3. The adjective "pragmatical" is to be used sparsely, close to its sense in [Morris, 1938], and only in connection with one or more of the terms "semantic role", "logical role", "problem situation", "coordinate system", "operation".

DEFINITION 3. An attempted pragmatization of a given theory is to be regarded as *successful* only if none of the power of the theory in its pre-pragmatized form is lost in the process.

3.4 States of activity

Below we shall have successful pragmatizations in mind. Furthermore, let us introduce a semiotical distinction between two (or more) *phases* or *states* of activity.

DEFINITION 4. By *State A* we shall indicate any mental or social state or activity of *communication*. By *State B* we shall indicate any state or activity of *preparing for communication* (by observing, experimenting, thinking, or writing without yet literally addressing any other persons). By *State C* we indicate a mental and social state of disinterest in (aversion or indifference to) communication.

This terminology is not intended to carry historical overtones. One state may yield to the next within seconds, or within millennia.

Among non-State-A oriented theory constructions one easily observes the existence of quite distinct philosophies and theory types: some favour *subjective description* of cognitive introspection or of emotions, others *objective description* of phenomena, persons and processes in the outer world, including the theoretician. We accordingly distinguish two non-A States, B and C, as the solitary user-theoretician does or does not pay attention to physical phenomena and individuals in an outer world. (For the purposes of reconstructing DL it is sufficient to distinguish two, A-states and the rest.)

State-A and State-B theory somehow or other contains *Multi-Role Description*, allowing for such category combinations (n-tuples) as

⟨Speaker, Hearer⟩
⟨Opponent, Proponent⟩
⟨Laboratory worker, Reporter, Theoretician, ...⟩
$\langle m^1, m^2, \ldots m^n \rangle$, with "$m$" for "measuring system"
$\langle c^1, c^2, \ldots, c^n \rangle$, with "$c$" for "coordinate system"

3.5 Weak and strong pragmatization

Finally we introduce, for reasons of historical analysis and meta-theory, the distinction between *weak and strong pragmatization:*

DEFINITION 5. By *strong pragmatization* we subsume theory transformations and processes that involve a shift of outlook from either State B or State C to State A.

By *weak* (or *pseudo*-) *pragmatization* we shall subsume processes that pay no attention to State A but are focussed towards a shift from State B to C (from *no-role* to *one-role* description).

For strong theory pragmatization we might choose the term *Naess transformation*, in honour of Naess' 1939 dialogical transformation of the semantics displayed in logical positivism (Naess [1956; 1992a; 1992b]). When, as some neo-Kantians (but certainly not Hintikka) do, the outer world is dropped altogether, one

can speak of a *solipsist transformation* of objectivist theory — though hardly of pragmatization (cf. the creationist-intuitionist form of mathematics).

3.6 The sources of logical (in)validity: problem-solving capacity and institutionalized conventions

Consider the notion of "self-evident" assumptions and statements, and the *postulate of self-evidence* (Beth [1959a; 1968]) in classical theories of deduction:

(Act 5) Relieve the philosophy of logic from the notion of "intuitive self-evidence"! Think instead in terms of the following possibilities of pragmatizing the notions of "validity" and "invalidity" for systems of basic principles (categories, formal$_3$ and formal$_2$ rules):

DEFINITION 6. Among the kinds of "logical validity" (or legality, or acceptability) of (systems of) basic *rules* and other principles for DL we recognize

- *problem-solving validity*: the system, rule or principle is seen to have *more problem-solving capacity* than all competing systems or principles in solving a given set of problems of a general logico-intellectual nature;

- *semi-conventional legality (in a certain circle of users of language)*: on account of a non-documented convention demonstrated in *very* common verbal behaviour and thought; or

- *institutional, or conventional legality*: the system, rule, definition or principle is "logical law" on account of a documented agreement (social contract) which is respected in certain circles, and which is based on this system's or principle's assumed problem-solving excellence.

By *logico-intellectual problems* are meant the wide problems of verbal representation connected with criticism and justification: how to handle non-affirmation; how to express relatedness among persons, phenomena, or things; how to deal with *concessions* pertaining to individual opponents; which categories should be chosen as logically fundamental. Such problems may be "solved" in various, not necessarily satisfactory manners. Given the logico-intellectual problems to be solved by some system or principle, "problem-solving validity" is a comparative concept.

The word "necessity" is too ontologically and metaphysically loaded to convey what is meant here. The same goes for "objective validity". Crawshay-Williams [1957, p. 224] therefore spoke of "empirical" or "methodological" necessity.[9]

We do not here go further into Beth's *invalidating models* (the method of semantic tableaux).

[9] More detailed definitions in [Barth and Krabbe, 1982, pp. 19–22].

4 MAKING MODERN LOGICS "DIALECTICAL"

An early discussion calculus — The adaptation problem (2) — Externalization of dialectics — Semiotical roles — Logical roles — An opponent's task differs from that of a proponent's, Def. 7 — Rhetorical, epistemological, and DL terminology — A series of "bones of contention", and unconditional defence of an intermediary thesis, Def. 8 — Invalidity pragmatized: from truth-conservation in all isomorphic$_2$ cases to having a winning strategy, Def. 9–10 — Empirical "validity", Def. 11.

4.1 An early discussion calculus

Naess has mainly used informal methods. In 1936 he wrote a paper where he formulated a "discussion calculus", meant to come as an addition to the calculi of the philosophers of the Vienna Circle, with whom he spent a year. They were, however, unwilling to see his recommendation as a necessary addition to their own calculi; as a consequence, he did not publish it. See his [1992a; 1992b]).

The discussion calculus was meant for use in those cases when the difference of opinion may be assumed to be caused by the vagueness of the thesis — a lack of precision. Then the question may rise: what exactly is the topic here? What is the initial formulation (thesis) actually meant to say? Vagueness may lead to *pseudo-disagreement*, but it can also provide *pseudo-agreement*, and this, too, may be undesirable in the long run.

[O]	[P]
I disagree	T_0
Well . . . at least, T_0 must be made more precise.	But why?

T_0 stands for the *initial thesis*. (Read from right to left.)

Naess was interested in the problems of procedure in such cases and formulated 6 basic rules. "Such a discussion logic is of great use in the discussion of political programmes."

4.2 Solving the adaptation problem (2)

The deductive frameworks for logical theory — axiomatic deduction, natural deduction, deductive tableaux — wrongly suggest that "logic" is essentially tied either to State B or to State C activities. Given True "premises", truth is guaranteed at the end of a number of "valid" steps. Falsity and invalidity are treated as mere "privatives" — nothing but absences, and nothing to speak of or count with. Neither of these methods can accommodate the categories of critical dialogue and "rational discussion".

The truth-table method and the method of semantic tableaux, operating with the category of Falsity on a par with Truth, are less suggestive of a solipsist basis for logic and rather invite objectivist associations.

We may well speak of the emancipation of the truth-falsity *distinction*, a development that owes much to both Popper and Tarski (and nothing to Wittgenstein and Gadamer). The next station in the dialogical development is the distinction of Opponent and Proponent. In fact, any "objectivistic"-semantical tableau (left) can, without loss, be reinterpreted as a dialogical tableau (right):

True	False		Opposition	Proponent
{Premises}	Concludendum	→	{ Concessions }	Thesis

(But not necessarily vice versa. It depends on the strength of the DL employed.)

DL can, if one so likes, be called "post-modern", in the sense that it is achieved by a *successful strong pragmatization* of "modern logic". This we now go on to show in some detail.

4.3 Externalization of dialectics

Above we pointed to uses of "dialectics" in ontology-related ideologies, such as interiorist idealism, Hegelian idealism, and Marxist–Leninist political theory.[10] In our century the classical use of the word "dialectical" relating it to dialogue and discussion is taken up again: by Naess [1936; 1992a; 1992b] and by Hamblin [1970]. In connection with what one could call the externalization of the "rational" world picture of earlier philosophy, described in [Hacking, 1975], let us recommend an "externalization of dialectics".

4.4 Semiotical roles

In our time everyone will agree that logic is in need of *initial distinctions of a semiotical nature* — at least one. A first distinction will be the semiotical distinction of *Hearer* and *Speaker*. Obviously, these are *roles*, usually to be taken now by one, now by another participant in a dialogue.

4.5 Logical/dialectical roles

A dialogical logic requires another distinction of roles in addition: that of *dialectical roles* (or logical roles, as we could also call them). This is Lorenzen's distinction between *Proponent* and *Opponent*, which presupposes that a *thesis* is in the picture, with some measure of disagreement as to its acceptability.

Proponent and Opponent, or P and O, for short, are roles, not persons or groups of persons. These roles are relations involving four terms: a person or well-defined *company of persons* [Crawshay-Williams, 1957, p. 10], a verbalized conflict, a

[10] For Lenin's logic see [Smit, 1996]. On Gramsci cf. [Finocchiaro, 1992].

system σ of a formal$_3$ dialectics, and a σ-based discussion among members of the said company, issuing from the conflict.

When, technically, a discussion is referred to as a "game", the debaters are often pertly referred to as "players" (about the game concepts involved see the paper [Dascal *et al.*, 1996]). So we might also say that Proponent and Opponent are roles the players take, not the players themselves.

The term "dialectical role" adopted here is taken from one of the several names that throughout the ages have been given — each with a historical variety of meanings — to the study and practice of general reasoning: *"dialectics"*, the term that was commonly used in several centuries, other names for logical studies being *"logica"* and, in Germany in eighteenth and nineteenth centuries, *"Vernunftlehre"* [Scholz, 1967]. In antiquity it carried overtones of an exchange of ideas between minds, by convincing arguments, a meaning which was lost during the nineteenth century.

There is of course nothing against one and the same person, party or machine taking on both roles simultaneously. This self-critical case, which was for some time emphasized in the Marxist regimes, is not without its value, but it ought not be the only one allowed.

4.6 Preparing for debate

Beth's method of invalidating examples, based on two truth values, was conceived some ten years earlier. Its nature emerges when one contemplates the the Truth-Falsity distinction from the point of view of the party in Opposition: "Suppose the concessions are true; the problem then is to see whether the thesis might *possibly* be false".

The emancipation of Falsity and the return to an oppositional point of view go together. Beth's method of Semantic Tableaux reflects a refutational, oppositional outlook, however without evoking the neo-Platonic notion of "Opposition in Being".

Beth's semantical tableaux may be regarded as a method of *preparing for discussion* (cf. Definition 4 above).

4.7 An opponent's task differs from that of a proponent's

DEFINITION 7.

a. *The Opponent's task* consists in trying to show that, although this Opponent has made certain concessions, a sceptical attitude towards the Proponent's initial thesis is advisable.

   This can be done by showing that the person in the role of Proponent can be made to exhaust his/her possibilities of refuting the Opponent's critical endeavours, and so has to give up. (Weaker or different concessions, a smarter

person in the role of Proponent, or a night's sleep may of course change the picture.)

b. *The Proponent's task* consists in trying to show that the position of the other party has become irreconcilable with this party's task and status *as a critic*, as shown by his/her moves in the discussion.

4.8 Rhetorical, epistemological, and DL terminology

Rhetoricians have things to tell of value for speakers who want to *convince* an audience, whether passive or not, and stand in a tradition that has collected impressive amounts of humanistic wisdom.

In DL the Proponent's task is to counter the Opponent's criticism of the thesis. The Proponent's position is a counter-counter-position. This is not at all the same as bringing about conviction, which may or may not result from the Proponent's success. In dialogue logic the terms "to convince" and "conviction" have no theoretical function.

What about the term "doubt"? Lorenzen's use of *Dubito?* as a symbol for the Opponent's challenge may at first bring to mind the epistemological deliberations of Descartes; these, however, are of a monological, interiorist nature, centring on "conviction" and "knowledge". In his *Descartes*, Gaukroger distinguishes between *epistemological doubt*, which questions "whether one's beliefs amount to knowledge", and *doxastic doubt*, which questions "whether one is entitled to hold the beliefs one has" [Gaukroger, 1995, p. 311]. It will be clear that Lorenzen's *Dubito?* is, in a strong sense, doxastic rather than epistemological.[11]

4.9 An interlocking series of "bones of contention". Unconditional defence of some intermediary thesis

When P's attempt at defence of T_0 is a statement, and that statement in its turn is challenged by the Opponent, that statement becomes a new thesis: the first intermediary thesis, T_1. If T_1 is defended by means of yet another statement which is in its turn attacked, it becomes T_2, and so on. At every moment the last intermediary thesis functions as the bone of contention until a new intermediary thesis is born.

In order to refute the Opponent's position the Proponent must, for some n, be able to give an *unconditional defence of* T_k (from attacks by the present opponent). Since T_k was introduced in defence of T_{k-1}, that statement now is defended, and so on, up to and including T_0 *(in a certain line of attack; see below)*.

What is to count as an unconditional defence? In some cases it will seem reasonable to say that the Proponent now has indeed shown the *untenability* of the

[11] A dialogue-logician's answer to Decartes' "Cogito, ergo sum" could be: "Dubito, ergo sum, vel es, vel est". Cf. also [Wiche, 1999] (possibly also Herlitzius, see above).

other party's position in the current discussion. In what cases? We suggest that DL should comprise, by agreement, the following "formal $_3$" rule, or definition:

DEFINITION 8.

a. When, at some stage, the Proponent (or a third party, whose judgement is, *by agreement*, to be respected by both parties) can show that the Opponent

- has just challenged a statement equivalent to one that he/she has personally stated earlier in the current branch of the argument;[12] or
- has just made a statement equivalent to one made by the Proponent uttered at an earlier point and which the Opponent then challenged (so that the Proponent ran into an obligation to defend it);
- [has conceded/presumed both a certain statement and its negation; or has uttered, as a statement or as an hypothesis, an acclaimed absurdity];[13]

or

- is unable to, on request, explain the *relevance* of his remark *to the thesis in question;* or
- has produced an abusive *argumentum ad hominem;* or
- has nothing more to question or to say in this line of attack, and cannot think of another one either; or
- has withdrawn from the discussion; or
- (this second list is not complete for all times)

then the Proponent *has unconditionally defended the intermediary thesis in question.*

b. When in a certain line of attack the Proponent has unconditionally defended T_0 against the Opponent's arguments, then the sub-discussion consisting in this line of attack is to be treated as *closed*.

c. When all sub-discussions defined by a line of attack are closed, the discussion as a whole is closed.

When a debater, be it Proponent or Opponent, abandons a discussion before it has come to an end, and no other rules intervene, he or she may be said to have "lost" it; and let us then say that the other party has "won".

Concerning arguments *ad hominem* and other speech acts that are unacceptable to the other party (e.g. remarks felt to be insulting the other party's culture) one could add:

[12] I.e., when the Proponent or someone else (etc.) can say: *"You said so yourself!"*.

[13] The bracketed conditions are not independent of the other ones. More details in [Barth and Krabbe, 1982].

d. If one of the debaters performs a not explicitly permitted speech act which reduces the other party's chances of carrying out its task, that party may honourably withdraw from the discussion without "losing" it.

For a number of cases of undesired speech acts one could choose to reformulate the last rule as follows:

> …then the party performing the speech act in question loses all its rights in the discussion as a whole.

Furthermore:

> As soon as the Proponent has unconditionally defended the thesis in question, the (intermediary) discussion should be regarded as ended.

No such rule could apply to an opponent who in a *simple* (one-thesis; see below) discussion does not have any thesis at all!

We assume, then, that the concessions (if any) function merely as material for use by the defence. This is the pragmatical arrangement of things that comes closest to what has in our century been called (modern) logic, *without being itself deductive*. Certainly one could conceive of forms of discussion in which each party makes concessions, the product of which is simultaneously presented as its thesis; but that would be another story, and a more complex one.

These rules will, in the long run, contribute to the dynamical character of the chosen system for rating results; i.e., it will contribute to new discussions and thereby to the *flux of opinion* in culture.

4.10 (In)Validity pragmatized

Criteria of validity in older renditions of the deductive method were formulated in terms of "logical axioms", somewhat later in terms of "premises", "rules of deduction" and "conclusion". The so-called conclusion "follows from" the premises "by rule so-and-so". There are no actors and no roles — or there is one actor, the Subject, as in creationist-intuitionist constructivism, without distinction of roles.

A non-deductivist definition of formal$_2$ validity preceded the dialogical definition by only a few years. This is the semantical concept of validity that is the basis of the Semantic Tableau method. The logical validity/invalidity of an argument lies in the absence/existence of formally$_2$ analogous counter-stories. It may be formulated as follows:

DEFINITION 9.

a. A counter-story to an argument from certain premises to a given conclusion is a story with the same number of premises, each one of the same syntactical form$_2$ as one of the initial premises, and with a conclusion of the same syntactical form as the said conclusion, in which each premise is *recognized as true*, whereas the conclusion of the counter-story is *recognized as false*.

b. The step from the premises to the desired conclusion is called "(formally$_2$) valid" if, and only if, *any attempt at finding or constructing such a counter-story is demonstrably doomed to failure*. [Beth's Criterion]

Now shift to conflicts of opinion. By *a simple conflict of avowed opinion* we shall understand a bundle of four things (see above):

a set of concessions (Con),

one initial thesis (T_0),

a party who (in the light of these concessions) takes the role of Proponent of the thesis,

one party — another, or the same — who takes the role of Opponent of the thesis.

It is to be understood that the thesis is a statement made (and not withdrawn) by the former party, and that the concessions are statements made or endorsed (and not withdrawn) by the latter.

Notice that neither "intentionality" nor any other interior feature is part of this quadruple. No assumption is made about the Proponent *believing* T_0 to be correct, or about O believing the concessions to be correct. We are not concerned with so-called propositional attitudes. We have only to do with *external* conflicts and the *externalized* aspects of conflicts and institutional aspects of critical dialogue: *statemental attitudes* (e.g. "commitments").[14] *To believe* is, primarily, a religious or political category which need not bear upon non-doctrinal discourse where "doubt" comes to the fore.

Whether we are looking at a DL that is entirely formal$_2$ or wholly or partly informal, if it is supposed to be a system of *logic* we shall have to consider what is to count as (in)validity, or else turn up a critical concept that can function in its place. In a strong pragmatization of deductive logic this seems an absolute requirement.

One dialogical criterion, based on formal$_2$ conditions, with the desired properties results by pragmatization of the semantical criterion.

In 1959 Lorenzen introduces into logic the game-theoretical notion of a "winning strategy". In some cases of T_0 and Con, it is possible to show that every opponent, however smart, who makes these concessions runs the risk that all his lines of attack will be closed. So it makes sense to formulate a definition of "validity" as follows:

DEFINITION 10. Given a system of dialectical rules that incorporates Definition 8, an addition of T_0 to the Opponent's avowed concessions is called (deductively) valid according to that system if and only if every opponent — however smart — who makes these concessions runs the risk that all his lines of attack will be closed.

[14] More in my [1982].

Or,

Given a system of dialectical rules that incorporates Definition 8, *T_0 is valid-in-the-light-of Con* (etc.).

4.11 Empirical "validity". The importance of time

DEFINITION 11. A thesis T_0 may be called empirically valid-on-the-background-of-Con-at-time-t if there is some party who in the role of Proponent has unconditionally defended T_0 against any party who has hitherto cared to act as Opponent on the background of Con.

The reference to time is crucial. At a later point in time the role of Opponent may be taken by someone who does succeed in the task of refuting T_0 in discussion with the same proponent, or with another one who is considered strong enough to take on the Proponent's task. Another possibility for change is to treat one of the concessions as a *thesis*, showing its untenability-in-debate vis-à-vis the following stable concessions ... (a new set).

As the reader can see, we fuse Popper's falsification perspective (and respect for discussion) with a firm denial of his hypothetico-deductivist outlook on the logic of accountability.

5 ACCOUNTABILITY

Accountability (1) — Critical moves and defence moves — Schematic rendering of the terminology — Accountability (2): Defensible interpretation and precization, Def. 12 — "I need an explanation of relevance" — Thematic discussion, Def. 13 — One or more theses? — Opponents' concessions — Maximum length of a contribution (stage, period) — In a thematic discussion rights and obligations depend on former remarks — A plurality of lines of critical attack on the Proponent's last statement — Accountability (3): As Proponent one is obliged to defend a challenged position — A plurality of lines of protective defence? — Refutation, Def. 14 — Defence, Def. 15

5.1 Accountability (1)

Here are some rules — informal_1 ones, like in other modern logics — that, when used as constitutive of a framework for DL, lead to a strong pragmatization of deductive logic, and that are so general that even informal logicians may be assumed to approve of them.

(Remember that already the modern framework of deductive logic assumes no forms_1 in the traditional sense, and of course neither does DL^*, the dialogical image of modern deductive logic.)

5.2 Critical moves and defensive moves. "Personal self defence"

There are two types of (verbal) move: moves of critical attack and defence moves against a given criticism.

Critical/attack moves are always directed towards a certain statement or a complete sentential component of a certain statement made by that party at some point in the discussion. A move criticizing a statement may be called a *challenge* to the party that made the statement *to defend* it, or a certain part of it.

(A remark on the terminology: as often as not, personal *self defence* is as well carried out by means of a *challenge*, or *critical move*, or *attack move*, as well as by *a justificatory* or *defence move*.)

Each criticism of a statement U (for utterance) might be called crit(U), but there are often several ways of criticizing or challenging a statement, and one may make lists of them. As theoreticians we may speak of the ith of the possibilities of challenging an utterance of a statement, U, as $\text{crit}_i(\text{U})$.

Notice that *a criticism or challenge of one of the other party's statements is often launched simply in order to defend some remark of one's own from (further) criticism*! There is nothing against it. On the contrary, to do so helps probe the tenability of the other party's position as a whole. In such a case the reader may dislike our use of the word "statement", but we shall use this word for all *declarative utterances made in a discussion*, irrespective of the utterer's beliefs. Our use of "statement" includes attitudes of complete disbelief, a hypothetical stance and firm belief.

Justificatory/defence moves are moves in defence of a certain statement of one's own, against a *certain criticism* of that statement.

There may be more than one way to do this too, even when the critical remark has already been made. *Any* possible defence of a statement U against $\text{crit}_i(\text{U})$ will be called $\text{def}(\text{U}, i)$, or def(U) for short.

5.3 Schematic rendering of the terminology

We sum up the terminology employed so far schematically (the letter "U" stands for "utterance of a statement in the given discussion"):

U	Possible criticism (challenge) of U:	Possible moves in defence of U against criticism $\text{crit}_i(\text{U})$:	
	$\text{crit}_i(\text{U})$	$\text{def}(\text{U}, i)$	
		Direct, protective:	Indirect, counter-active:
		$\text{prot}_j(\text{U}, i)$	CA

(Notice that we have not yet applied these terms to the specific roles of Proponent and Opponent.)

The criticism $\mathrm{crit}_i(\mathrm{U})$, essentially a request for a justification or an elucidation of some kind or other, may itself be a question as well as a statement. However, provided the critic has remained within the exchange conventions, $\mathrm{crit}_i(\mathrm{U})$ may be criticized only if it is itself a statement.

An utterance $\mathrm{crit}_i(\mathrm{U})$ may be called a *counter-argument* to U in the sense of *Pro-aut-contra* analysis only when it has the form of a declarative sentence.

Now suppose $\mathrm{crit}_i(\mathrm{U})$ is a statement. Then a defence $\mathrm{prot}_j(\mathrm{U}, i)$ of a statement, may be called an argument *contra* $\mathrm{crit}_i(\mathrm{U})$ and a *pro-contra*-argument for U. DL may therefore be seen as a development of *pro-aut-contra* analysis. In DL arguments *pro* a certain declarative utterance U are studied as functions not only of U but also of the *challenge (criticism)* that has been advanced against U.

In the following we shall formulate a number of rules for what will count as an acceptable challenge of some declarative utterance U, and for what will count as an acceptable defence of U against a given (acceptable) challenge.

What to do in case of *not* acceptable challenges is a matter that will have to be dealt with separately.

5.4 Accountability (2): Defensible interpretation and precization

Naess' publications contain some of the strongest pragmatizations of topics of logical interest in advance of DL.[15] They assume a dialogical setting. In this setting Naess gives a two-step definition of preciseness, more precisely: of *more precise than*. We shall try to express the dialogical intention as clearly as possible:

DEFINITION 12.

a. A listener can "reasonably", i.e., with high probability that many other people will agree, take a formulation F' as an *defensible interpretation of* another formulation, F, if and only if there are one or more connections in which they both can mean the same.

F' could also be called a *likely* interpretation of F, but this term, too, leaves out the dialogical vista. Worse still, in such a case F' has often been called a "reasonable" interpretation of F, a vernacular liable to produce reflections on monological "rationality" rather than on dialogical efficiency. This is not taken from Descartes. Notice that we are talking about linguistic entities.

b. A formulation, say: F'', *is more precise than* another formulation, F, when F'' *excludes* some *defensible interpretations* F' of F, but not *vice versa*.[16]

The (re)formulation F'' may then be called a *precization of* F and F a *deprecization* of F''.

[15] Cf. Naess [1953; 1981; 1992a; 1992b].
[16] Reformulated from [Naess, 1981].

Clearly, of two formulations only one can be more precise than another before *the same audience or readership*. One might insert a variable for the audience/readership into this definition.

As defined by a horizontal "dialogue strip":

$$\text{U} \quad \text{Prec?} \quad \text{U}'$$

In many cases of unclear sentences precization can be reached by means of a (lambda-)analysis of syntactical form, showing their ambiguity. In other cases a clarification of an epistemological nature may be more helpful.

In courses of practical logic what is here presented as Definition 12 may reasonably be dealt with first, since critical discussions will often require an initial exchange of precizations.

5.5 Relevance

Similarly, a debater may ask for some information on the *relevance* of an utterance the other party has made. There is no widely accepted definition of "concept C is relevant". We recommend to look instead for a definition of

C' is more relevant than C to T_j, the (sub)thesis in question.

In any case, in discussions with benevolent and obliging discussants it is often a good idea to ask for a *more thesis-relevant* conceptualization. The dialogue strip is simply:

$$\text{C} \quad \text{Rel. to T}_j\text{?} \quad \text{C}'$$

5.6 Thematic discussion (1)

We give this term an open definition:

DEFINITION 13. By a *thematic discussion* — the term is intended to replace the old "rational discussion" — we shall understand any discussion in which, at any stage, the participants' rights and duties are *functions of* (*among other things*) *the basic theme*, here identified dialogically as the initial thesis of the discussion. More often than not a discussion is carried out between two parties, who speak alternately, but no dogma dictates that two is the limit.

In any discussion there is — according to the lexical meaning of "discussion" — more than one "party", i.e., there are two or more *logical*, or *dialectical, roles* to be taken.

A dialectical theory distinguishing a third party ("the Judge") was suggested by Herlitzius at the International Congress of Logic, Methodology and the Philosophy of Science in 1967, but this approach does not seem to have been developed any further; besides, the notion of a judge seems unfortunate.

Recently Wiche proposed the same arrangement (formal$_3$ rule) for informal$_2$ logic [Wiche, 1999]. (He speaks in terms of a "leader" of the discussion.)

Like Naess' several innovations and recommendations, this may be a good idea also for the informal$_2$ extension (stage setting) of formal$_2$ DL.

The parties in DL can be (emotional, political) antagonists, but they can also be mutual supporters in testing an hypothesis.[17]

Notice also that one person may discuss with him- or herself and then has to play both a *positing role* and a *critical role*, as two parties.

5.7 *One or more theses?*

We ask the reader's special attention for the following.

1. In a *simple* Proponent-Opponent discussion we assume one thesis (hence no "antithesis". The expression, much favoured by Hegelians, reflects the semiotics of their concept of "opposition". This is one central differences between dialogical and idealist "dialectics").

2. The one and only initial thesis, T_0, of a simple discussion is the dialogical equivalent of what is called the conclusion — the "concludendum" — in a deduction. It may be felt that even a non-bellicose discussion may have more than one initial thesis, some stated by one party and some by another, some by a third .... In practice this is often the case, which makes for considerable confusion. (The opponent of one thesis may be the proponent of another one, which may be called an "antithesis" if, and only if, the two (or more) theses are contraries.)

   It may seem harmless — let us put this tentatively — to approach such discussions as just an untidy superposition of simultaneous one-thesis discussions, and to recommend that one-thesis components be carried our consecutively, for the sake of clarity and efficiency.

   However, one should not exclude the interesting possibility of *interference phenomena* between simultaneous one-thesis discussions.

3. A related important problem sphere is what one calls non-monotonic logic, i.e., a logic which pays attention to new premises that come into the picture in the course of an ongoing debate. I am inclined to think that interferences should be approached from there. Van Hoof puts non-monotonic logic into (semi) dialogical form, introducing a new operator for non-absolute negation [Hoof, 1995].

4. *Caveats*: The party in opposition is the opponent of T_0 — and not of the person in the role of proponent. This should be clear enough. To ask whether or not the party in the role of Opponent is *thereby* an "opponent of party P" is then merely a dull semantical question.

[17] Cf. Gilbert [1994, n. 39].

Even in a complex practical situation where more than one party claiming to have an initial theses, each of these parties may be understood to retain his/her role of Proponent or of Opponent *of this thesis* as long as that particular thesis is under debate.

5.8 Opponent's concessions

In general, some opinions will be known to be accepted — at least for the occasion — by the Opponent of T_0; these will be called O's *concessions*, and the class of these concessions will be called Con.

The opponent's concessions may be known to P in advance or elicited in the course of the discussion.

In DL* one may assume that Con is given in advance. In dialogical non-monotonic logic this assumption is dropped (Van Hoof, above).

5.9 Maximum length of a contribution (stage, period)

The maximum length of a contribution (stage, period) in the dialogue can be determined in advance, by setting an upper limit either to the time used by one speaker at a stretch or to the number of sentences or of words used.

(In DL* we shall limit the number of sentences per contribution to 1, because this makes for the clearest way of solving the adaptation problem and of pragmatizing modern logic. Notice that any number of statements can be joined together with "and" so as to form one statement, in one sentence. This means that T_0 may consist of a whole theory, understood as one long conjunction of sentences. Hence this "limitation" is a not very important point, and in no way restricts other applications of the framework than the construction of DL*; it can just as well be adopted for a DL of any kind.)

5.10 Thematic discussion (2). Rights and obligations depend on former remarks in the dialogue

A *critical dialogue* (discussion) develops as a sequence of stages. These may be characterized by *sequents* — ordered pairs of the type $\langle \mathrm{Con}_k, \mathrm{T}_k \rangle$, which identifies the kth stage. Stage 0 is $\langle \mathrm{Con}_0, \mathrm{T}_0 \rangle$, and so on.

It is recommended to debaters to adopt the following systematizing vernacular: "Con_k comprises all the Opponent's *statements* (not the questions!) up to and including stage k", in other words, that Con be understood as cumulative.[18] Excepting for the initial thesis T_0, T_k may be called the *subthesis at stage k*.[19] The

[18] In 1978 Woods and Walton exploited the idea of a non-cumulative logic, closely followed by Mackenzie in 1979. A definition of a hidden traditional principle of cumulation of agreement may be found in [Barth and Krabbe, 1982, 243ff.].

[19] In [Barth and Krabbe, 1978] the term "local thesis" is used.

rights and duties of Proponent and Opponent at stage k of the dialogue can be described as *functions of the situation at that stage*. Both parties ought to keep this in mind. Thereby one ensures that the original "theme" of the discussion is kept constant through the whole debate.

This is the essence of a *thematic discussion*.

5.11 A plurality of lines of critical attack on the Proponent's last statement

An opponent may be faced with a plurality of promising and acceptable ways of challenging the intermediary thesis of the moment. Suppose this is the case at some stage k, the intermediary thesis being T_k. We abbreviate *criticism of* T_k as crit(T_k), and the various possibilities of criticism as crit$_1(\mathrm{T}_k)$, crit$_2(\mathrm{T}_k)$, and so on. (In practical logic, outside mathematics, we may assume the number of possibilities to be finite.) This may be illustrated by a *strategy tableau*:

$\vdots$		$\vdots$	
Con$_k$		T_k	
1	2 (etc.)	1	2 (etc.)
crit$_1(\mathrm{T}_k)$	crit$_2(\mathrm{T}_k)$		

We recommend that the chosen rules for any DL allow the Opponent to deal with each of them, provided there is time enough, and, *for clarity's sake, and for efficiency's sake* (see below, Section 6 and following), provided they are dealt with *separately*.

The subdiscussions could — if both parties agree — be dealt with simultaneously, which is often the case in practice; but this usually makes it much harder to keep track of what is going on.

Since the Proponent's justification always is a justification of the Proponent's position "as against the given critical attack", the Opponent's choice is, in principle, crucial for the rest of the discussion issuing from it.

Suppose the Proponent's defence is a success. This is not to say that the Opponent could not have refuted the Proponent's position through a different attack on T_k. If time allows, the Opponent may open several "lines of criticism" of T_k, one after another, each line of criticism amounting to a subdiscussion of the discussion as a whole. After each such subdiscussion that has been "won" by the Proponent the parties return to T_k, the Opponent may chose to attack it in a different manner; this initiates a new line of criticism, and so on. The Opponent has succeeded in his/her goal of refuting T_k, and thereby the initial thesis, *as soon as the Proponent cannot find anything more to say* (see below).

5.12 Accountability (3)

As Proponent, one is obliged to defend a challenged statement, at any stage of a discussion. Any critical attack made by the Opponent creates a new intermediary thesis. The Proponent has thereby incurred *an obligation to defend that new intermediary thesis* (T_k), against this attack. (The initial position is defended if, and only if, every intermediary thesis has been defended.)

One can do this protectively or counteractively. The latter will say that one forgets about protecting T_k for the moment and concentrates instead on *probing the Opponent's position*. Does not the Opponent, who after all has made certain concessions, in challenging this intermediary thesis, involve him- or herself in an impossible situation?

As long as this probing can be undertaken without the Proponent uttering anything else than critical questions, the Opponent's retort cannot produce a new intermediary thesis "overshadowing" the present one. That is to say, as long as no rule limiting the number of the Proponent's counteractive questions interferes (one could have such a rule). Besides, the Proponent can return to this probing activity if so happens.

This means that we can concentrate on situations in which the Proponent has a choice of ways of defending T_k *protectively*.

5.13 A plurality of lines of protective defence?

The Proponent will often find that several moves of direct defence of an attacked remark are open, each of which *might* lead to a finalized defence. This happens for instance when the Proponent has said something beginning with "At least one of them ..."; obviously the Opponent may ask: "Who?" The Proponent can, if he/she so pleases, answer by saying "John ..." but also "Ann ..." or "George ...", etc. Other examples are moves that are authorized by the dialogical definitions for "logical particles", which will be given below.

Each possible move can in principle be used to embark upon a new subdiscussion. The Proponent may be able to choose between several "lines of direct defence".

crit$_i(\mathrm{T}_k)$		T_k	
1	2 (etc.)	1	2 (etc.)
		prot$_1(\mathrm{T}_k)$	prot$_2(\mathrm{T}_k)$

(Or, for short, "p_1", "p_2", and so on; but remember that the special line of criticism of T_k within which the Proponent's choice takes place, may be one among many! In order to make this clear we could write "prot$_j(\mathrm{T}_k, i)$".)

Some would see a problem in the fact that a proponent can lose his/her first attempt to defend T_k protectively — viz. by saying def_1 — and may come to regret not having said def_2 instead. For if the Opponent attacks def_1 so that def_1

becomes the intermediary thesis, the Proponent may seem to have lost the opportunity of ever making use of def_2 as an answer. So debaters have to make up their minds about this question:

Should the Proponent be allowed to return to a former intermediary thesis T_k*, and take up the defence of his/her position at that point?*

Some may say (this is the liberal view): "OK, we still have some time, so why don't you give it a try". This demonstrably corresponds to the use of two truth-values *true, false*.

Others say (this is the stricter view): "No, if you had the audacity to utter T_k, it ought to mean that you were conscious of reasons for maintaining that belief, and if so, you would have introduced one of your better reasons — e.g., def_2 — right away. To withhold a good reason, if it be one, and come up with it later on can only mean that you did not really know what you were talking about in the first place: you did not possess a recipe with which to construct the answer." This stricter view corresponds to the outlook of Dialogical-pragmatist constructivism (Section 2.6).

The mathematical difference between two-valuedness and constructivism is based on problems concerning vague notions like "infinite sets (lots) of numbers". When the problem under discussion has nothing to do with "infinite" lots of individuals, a distinction between the two views is less important, if at all. In practical life questions of "infinite" numbers do not come up.

There may be cases, however, when one may have fair reasons for maintaining a more complex thesis (a former intermediary thesis) involving T_k, without knowing whether this T_k is itself correct (true). It therefore seems fair to allow the Proponent the same freedom to try out several *lines of defence* of T_k, in answer to one of the Opponent's attacks.

5.14 Refutation

In a *simple discussion* the person or party in the role of Opponent has, in this role, no thesis to defend. In this sense the positions of Proponent and Opponent are asymmetrical. However, the Opponent, too, has something to "defend", though it is of another nature: namely, his or her own position as opponent of the thesis in question. In the course of the discussion the Opponent usually makes several challenging but also puts in some utterances that do have declarative form, the form of a statement (however tentative and ad hoc they may be) — for example, the concessions. When the Proponent challenges one of the Opponent's statements, it is done in defence of the thesis, with the aim of probing the Opponent's total position in the debate.

Dialogue logic makes one aware that to have "refuted a thesis" is a notion that must be related to proponents as well as to opponents. To have "refuted" a thesis against a certain proponent only means *to have countered the defence of that thesis, as against a given set of lines of attack, and as carried out by that specific proponent*. Hence:

DEFINITION 14. For O to have *satisfactorily invalidated P's position* it suffices that O has won — in the sense that P has had to give up — in one line of attack (subdiscussion opened by O him/herself).

In that line of attack O must have won in each and every line of defence opened by P.

If O fails to "refute" T_k (on the basis of Con_k) in *one line of attack* (a subdiscussion starting from $\mathrm{crit}_1(\mathrm{T}_k)$, then he or she may, if time allows (a question DL does not pronounce upon), attempt to do so in a new line of attack (a subdiscussion issuing from $\mathrm{crit}_2(\mathrm{T}_k)$), and so on.

In DL, it is not the case that as criticism of a thesis, anything goes.

5.15 Defence

Obviously, to defend a thesis against a certain opponent means to refute that opponent's attempt at refutation of that thesis:

DEFINITION 15. P has *satisfactorily defended a thesis against O's opposition* only when he/she has countered each and every one of O's lines of attack. I.e., P must win each and every subdiscussion opened by this opponent.

In some cases the remark that O wants to challenge is such that O can only express a general doubt, a move that opens only one subdiscussion. However, in such a case P will often have the opportunity of choosing between several lines of defence. This is the case for instance when the (sub)thesis explicitly concerns some, but not necessarily all individuals of a certain class. If the thesis is *Some Golliwogs are criminal* the Opponent can only say: "Who?", "Name someone?" or something similar; it is then sufficient when P names one Golliwog and asserts the criminal behaviour of that individual. Another such case is when the thesis is an inclusive "or" (p, q or both); P can choose to defend *p* or to defend *q*, and need not defend both.

> *P has won a (sub)discussion opened by O as soon as he/she has won one line of defence in this (sub)discussion.*

If P fails to defend T_k (on the basis of Con_k) *in one line of defence* (a subdiscussion following a critical attack $\mathrm{crit}_i(\mathrm{T}_k)$, then he or she may, if time allows, attempt to do so by invalidating O's position in a new line of defence, and so on.

Each one of the — perhaps many — subdiscussions opened by O may occasion several subdiscussions to be opened by P, so that the total discussion branches out in two manners simultaneously.[20]

It is easier to take in the situation if we agree to say that each subdiscussion, though perhaps *opened* at a later moment, *begins right at the top*, with T_0, the Opponent's concessions and the Opponent's original challenge of T_0. I shall assume that this agreement is made. Now notice that

[20]See examples in [Barth and Krabbe, 1982, 133ff], and in [Lorenzen and Lorenz, 1978].

> *the Proponent may open lines of defence (against some attack on the thesis or sub-thesis) by means of a counter-attack: crit*$_i(\mathrm{U})$

on one of the Opponent's (initial or later) concessions, U, in the subdiscussion in question. There is, then, a crucial distinction between "line of defence" and "discussion line opening with a protective move".

There is a great difference between distinct lines of critical attack, *opened by the Opponent*, and distinct lines of justificatory defence, *opened by the Proponent*. Writers on problems of argumentation often disregard the importance of this distinction.

As examples of "defining strips" for logical constants, take those for *and* and the inclusive *or:*

U	crit$_i(\mathrm{U})$	prot$_j(\mathrm{U}, i)$	CA
$p\&q$	$p?, q?$	p, q	
$p \vee q$	$(p \vee q)?$	$p; q$	
$p \rightarrow q$	p	q	
$\neg p$	p	-	

The question marks may be read: "How do you defend that?" In the first case, the one who uttered U, a conjunction, must be prepared to be challenged both as to p and as to q; in each case he or she may choose between a counter-attack on the other party's position and a protective defence of the attacked part of the said conjunction. In the second case U is a much weaker statement and a critic can do no more than express doubt, or scepticism; here the party who uttered it is free to choose between a protective defence of p, or of q, or a counter-attack on the other party's position. A negation may be challenged by stating the corresponding affirmation; there is no other way of defending ones position than to look for a counter-attack.

The sentence constructors are now defined in terms of rights and duties pertaining to the dialectical roles, with no reference to truth values or axioms. The rules for the "quantifiers" (as they are still called) will be introduced in the next section.

III. Out of the fly-bottle

6 LIMBO 1: CHOICE-OF-INDIVIDUAL-CASE RIGHTORS

A world of impractical entities — Example 1: Arbitrary/fluctuating individuals — Example 2: Infinitesimals — "Quantifiers?" Role-related rights! — Dialogical analysis of Weierstrass' definition — Example 3: Definite-individual descriptors — So far: First-order pragmatizations

6.1 A world of impractical entities

"Chimeras" will be our term for the impractical notions of impractical entities that make discussion practically impossible. In any culture there are more chimeras than we can deal with here. In our introduction we mentioned several of them, from Impetus to the Bearers of Logos. All hamper effective discussion.

Many of them have been eliminated, some by moves towards a dialogical logic. In this section we shall discuss three famous examples, in chronlogical order: "arbitrary individuals", "infinitesimal entities (or numbers)", and descriptions of definite individuals, such as "the present King of France".

EXAMPLE 1 ("Arbitrary/fluctuating individuals"). Take, for a start, the category "arbitrary individual". The terms that are used to deal with them are called "arbitrary terms". What is the syntactic category of such terms? In order to formulate dialogical rules for discourse involving them we need a clear answer to this question. In many older texts no distinction is made between the concept of *an arbitrary M* and the concept of *being an M*. (Such a distinction requires something like a categorial grammar, and can probably not be made without recourse to the concept of *function*s — semantical, syntactical and cognitive.)

In our century this ultra-vague notion, so closely connected with racist, sexist or other *typology*, is rightly condemned, but not clarified. More and more "an arbitrary *individual*" did give way to "an arbitrarily *selected* individual term" and then to "an arbitrary *choice* (of an individual or term)"; nonetheless the word "arbitrary" remained embarrassingly opaque. It served different ends: it was used in modern deductive logic to formulate both the rule of Existential Instantiation (EI, pertaining to the elimination of the operator *there is* by bringing in an individual case) and the rule of Universal Generalization (UG, pertaining to the insertion of the *all*-operator where only an individual case had been mentioned). In both cases, the learner is taught, for the rule to be "valid" the individual-term must be "arbitrary" or "arbitrarily chosen" — whatever that may mean. (Notice that in the first case this requirement comes in before, in the second case after the individual-term has been chosen!) But what is it to "arbitrarily choose" a triangle among a sea of triangles?

No less important: what is it, and under which conditions, to "arbitrarily" choose a human being a from a class of humans? At mid-century, after much discussion, a formulation was reached that would satisfy mathematicians and taught to students of deductive logic, but no one is likely to say that logic had now reached a stage adapted to practical use or to a logical scrutiny of discourse outside of mathematics.

Around 1907 a romantic notion of the "free unfolding" or "free arbitrariness", accompanied by a notion of "free choice", entered the scene of mathematics, as elements of the creationist-intuitionist philosophy inspired by St. Augustine as well as by German idealism. Here "free" goes together with "independent of the Truth–Falsity distinction", at the same time as exchange of thought and discussion

are belittled; mathematics is now the life of the *idealized Free (Creative) Subject*. This mathematics of the one, lonely Subject is an excessively proud one: "Like Kant, Brouwer founded a-priori mathematics on some ability of the individual mind *to grasp reality directly*" [Stigt, 1990, 180]. St. Augustine would have rejoiced.

EXAMPLE 2 ('Infinitesimal (beings, numbers)"). A sensational pragmatization of the mathematics of dynamics and change was obtained when Weierstrass replaced the notion of "(fixed) infinitesimal number" with a new concept, that of the *limit of a series of differential quotients* (a change in the value of some variable divided by a change in the value of some dependent variable). When the value change in the first variable is made smaller and smaller, the quotient S_n between the *n*-th change in the first variable and the *n*-th change in the dependent variable sometimes does reach a limit, S. In other words, a limit may be reached when the number, *n*, of differential quotients S_n "grows and grows".

Weierstrass defined the meaning of a statement about such a limit in the following terms: "*For every* ε, *there is a number* N (depending on ε) such that for all numbers *n* larger than this N, the difference of S_n and the limit, *S*, is smaller than that ε".

Briefly:

$$(\forall\varepsilon)(\exists N)(\forall n)[n > N \rightarrow /S - S_n/ < \varepsilon].$$

For some readers the latter, symbolic presentation of his definition is *more precise than — is a precization of* — the same definition as given in ordinary language, but not for all. Even on the symbolic reading the definition is usually not felt to be intuitively clear.[21]

6.2 *"Quantifiers?" Role-related rights!*

The process of pragmatization is not complete until the rules answer the question: *Who* is to select (the name of) an individual?

By "choice-of-individual rightors" we are referring to the seemingly simple words "every" (everyone, every individual) and "some". Let us call them *individual-rightors*, for short.

Using abbreviations introduced above one can formulate rules corresponding to common usage as follows:

U	$\text{crit}_i(\text{U})$	$\text{prot}_j(\text{U}, i)$
Everyone is P	a?	a is P
Some are P	?	a is P

[21] Cf. [Davis and Hersh, 1972], which I criticized in [Barth, 1982].

Or, in the (wider) modern terminology of syntactic/semantic functions:

$(\forall x)Fx$	$a?$	Fa
$(\exists x)Fx$	$?$	Fa

6.3 *Lines of criticism ("attack")*

Above we dealt with pluralities of possible lines of "attack" and "defence" in a general way. From the defining strips above in it will be clear that the rightor "For all ..." expressed by one party offers such a plurality of possibilities to the other party.

6.4 *Effective discussions*

It makes for effective discussions if both parties try to simplify the statements involved as far as possible. In practice one often does this (though the purpose of efficiency may not always be the strongest one among the purposes one is aware of in practice).

One type of simplification consists in probing the tenability-in-debate in *select individual cases*.

In case of an utterance using one of the linguistic forms *Every individual is* F or *All individuals are* F, the right to choose an individual, a, falls to the critic: *"I want a justification/defence of this as concerning the individual* a".

Thereby the party who uttered *Every individual is F* (as subthesis T_k) becomes committed to the simpler statement a *is F* (subthesis T_{k+1}), and has to defend it when challenged — see Accountability (3).

An utterance of the linguistic form *Some individual/s is/are* P (or equivalent) gives a critic no such right. The only way of expressing one's wish to explore its *tenability in debate* is to say so — briefly: "?". (Lorenzen here spoke in terms of doubt, "*dubito*".)

Traditionally, words like "all" and "some" have misleadingly been called "quantifiers", in homage to a metaphysical distinction between "quality" and "quantity". We recommend to not accept this highly misleading term, "quantifier", and to express this in our symbolism:

U	$\text{crit}_i(\text{U})$	$\text{prot}_k(\text{U}, i)$	
$x^c Fx$	$a?$	Fa	a chosen by *c*ritic
$x^s Fx$	$?$	Fa	a chosen by the party who *s*tated U

Or, in terms of that party:

$x^{you} Fx$	$a?$	Fa	a may be chosen by you
$x^I Fx$	$?$	Fa	a to be chosen by myself.

(Observe that in the course of a discussion, the Proponent of the initial thesis as well as the Opponent will sometimes act as speaker, at other moments as critic;

in other words, that all these rules are symmetrical in relation to the Proponent-Opponent distinction.)

6.5 Dialogical analysis of Weierstrass' pragmatization

However, a further pragmatization of Weierstrass' pragmatical definition in terms of dialectical roles may help. For (ii) can now be read: $\varepsilon^c N^s n^c$, where superscript c indicates the critic's right to choose a tolerable value of ε and superscript s the speaker's own right to choose a lowest value of N:

"*You* may choose an error (ε) that you will tolerate as to the practical deviation from the S in my statement; then *I* am prepared to provide an ordinal N such that, whichever ordinal n *you* may subsequently choose, *I* am prepared to defend that the difference $/S - S_n/$ is always smaller than the ε you say you can tolerate, provided $n > N$."

Now the sequence of operators in Weierstrass' definition has in fact become one of *I*'s and *you*'s in a certain order, indicating that choices be made in that order:

$$\varepsilon^{you} N^I n^{you} : \text{ if } n > N, \text{ then } /S - S_n/ < \varepsilon.$$

Note, again, that the person who utters a limit statement $\lim S_n = S$ (*when* n *is increased*), and who, on request, offers $\varepsilon^c N^s n^c \ldots$ as a precization, may well be the Opponent in the discussion! Let us repeat:

1. For an utterance dominated by a universal operator: $(\forall x)Fx$, any *critical listener* who participates in the discussion may select a value of x, say x^c, and require a justification (defence) of Fx^c.

   x^c has been called either an "arbitrary" or "arbitrarily selected" individual.

2. For a sentence dominated by an existential quantifier, $(\exists x)Fx$, the rule is: the *speaker* may select a value of x, say x^s, and defend Fx^s, if challenged.

But x^s, too, was called "arbitrary" or an "arbitrarily selected individual"! That is to say, "arbitrary" or "arbitrarily selected" was used so as to cover both cases indiscriminately — whoever did the selection (or, in whichever logico-semantical role the reasoning person, if alone, did the selection). However, having now recognized the logical importance of the two roles — which no one had done before 1960 — we no longer seem to have a job for the term "arbitrary", which appears as a hopelessly uninformative old ontological fog. A vague ontological category has been replaced by a pragmatical distinction.

In terms of Aristotle's variables S ("subject term"), P ("predicate term" — do not confuse this use of "P" with the use of "P" for "proponent"!) — and M ("middle term"):

To "arbitrarily choose" an individual with the property S given a universal premise *All* S*s are* P, is not the same as "arbitrarily" choosing an individual with the properties S and P in such a way that from *this* S *is* P one can conclude that

All Ss are P. Nor is it the same thing as "arbitrarily" choosing an individual with the property S on the strength of a premise *Some S is P*. And yet the locution "arbitrarily" is used in all these cases! But all these uses of "arbitrary" can be missed.

Due to Weierstrass' pragmatization of the (differential and integral) Calculus, as it is now called, the more general metaphysical notion of "infinitesimal being" suffered degradation as well. Its mathematical uses, being clearly superfluous, could no longer endorse its value.

Did the notion of "infinitesimal being" disappear? Not in the nineteenth century, for too much speculation was made to depend on it. Idealist philosophers in particular attempted "to interpret the differential as having an intensive quality resembling the potentiality of Aristotle, the *conatus* of Hobbes, or the inertia of modern science" [Boyer, 1959, 305]. More recently it has been shown, on an unpragmatical model-theoretical basis, that it is feasible (though not necessary) to go on with non-primitive artifacts called "infinitesimals", defined in that model. The model-theoretical conventions that make it possible to postulate "infinitesimal" numbers without running into contradictions will appear thrilling to some mathematicians, to others they are just superfluous. Thanks to Weierstrass, Beth and Lorenzen we possess a reformulation of the logic of movement and change that invokes no impractical chimeras. A successful pragmatization.

EXAMPLE 3 (Definite-individual descriptors). Now take statements with a description defining an individual, such as "The President is bald". Here the rightors come in another order and have other scopes. Given Bertrand Russell's analysis of the logic of such definite descriptions of individuals, and given the defining "strips" for individual-rightors, the further pragmatization is now easy: *Ask for a "precization"!*

What Russell's "theory" does is to offer such a precization. I.e., challenge any statement of the *form*$_2$.

The (one and only individual) F is G,

or in international symbols:

$G(\iota x.Fx)$,

by the challenge *Prec?* Direct defence move:

$(\exists x)[Fx\&Gx\&(\forall y)(Fy \rightarrow y = x)]$, and so forth.

In terms of our choice-of-individual rightors:

x^I : *x is President of France and x is bald (and, furthermore, if y^{you} is President of France, then y^{you} is identical with x^I).*

From here the discussion moves further as described by the strips for these rightors.

So far we have found the solution through first-order pragmatization. On our two-role reinterpretation of "all" and "some", it becomes remarkably easier to lift the veil of mystery from some particularly opaque terms that have mystified authors and readers, teachers and students since time immemorial. Given the strips for the individual–rightors, the expressions "individuum vagum", "indefinite/floating individual", "arbitrary individual", "arbitrary choice", have no additional job to do in our minds and none in language. They may be relegated to the same museum that accommodates "impetus" and "ether" (a presumed *carrier* of electromagnetic waves).

7 LIMBO 2: RIGHTORS OF HIGHER ORDER

Chimeras of the upper worlds, and second-order pragmatization, Def. 16, Def. 17 — Intentionality and the subject-object distinction

7.1 Chimeras of the upper world, and second-order pragmatization

The language of "being", "substance", "essence" (German "*Wesen*"), "form" involves so-called "generic" terms (descriptions) and statements.

Such language does not easily make for effective dialogue. But how do we recognize it? Let us say:

DEFINITION 16. An utterance *The* (or *a*, or *an*) X *is* Y has, for the person NN, in a certain context, a truly "generic" sense if and only if NN does not, in this context, regard it as fully synonymous with some statistical result or assumption, nor with any one of the following: *Every* X *is* Y*; Some* X *is* Y*; Most* X*s are* Y*; The one and only individual* X *is* Y.

(Notice that we do not refer to arbitrary or floating individuals here either!)

We said in the Introduction that since triangles possess no DNA, not all "generic" terms can be understood the same way. This definition covers all of them:

DEFINITION 17. From a dialogical point of view, any truly "generic" locution has a *message meaning* of the following form: "*There is* some principle, plan, recipe, metaphysical system, "middle term" (Aristotle), theological dictum, natural law, juridical law, axiom system, genome, scientific theory, or other background conviction, or some social institution, M, such that *on the strength of* M, each and every individual X is Y, whatever it may look like in the phenomenal world."

The situation is even more involved, since many uses — historical or contemporary — of generic sentences contain, overtly or tacitly, a notion of potentiality: "The (a, an) X is potentially Y". However, *outside physics an ontological notion of a "potential" has no clear dialogical meaning-in-critical-use*.

In other words, the dialogical sense of a truly "generic" sentence is this:

$$(\exists M) \longrightarrow_M \text{ Every } X \text{ is } Y,$$

that is to say

$$(M^I) \longrightarrow_M \text{ Every } X \text{ is } Y.$$

In this Limbo pragmatization by means of first-order rightors is not possible. We had to avail ourselves of a second-order one, $\exists M$.

The world at large has confounded "generic" with first-order "universal" as well as with first-order "existential" sentence forms, thus making effective discussion impossible.

SUGGESTION. An argument should be deemed *dialogically in*valid if (but not only if) it contains syntactic elements for which no dialogical strips corresponding to or sharpening the usage can be formulated.

This will usually be the case whenever the syntactical category of the non-referring particles employed is not clear.

From a dialogical point of view the problem is this: can a participant in a discussion challenge the other party's generic locutions? If so, can the former party defend that locution against that challenge?

On our definition this is the dialogical solution:

When addressed with a generic utterance, ask the speaker for the M-value on which the utterance is based!

Writing down the defining strip is now easy; for many users of language, giving a defensive reply will be hard.

The approach to the logic of generics by Hoepelmann and Van Hoof [Hoepelman and van Hoof, 1993; Hoof, 1995]), though probably adequate for the linguistic intentions of (semi)modern users of language, in my opinion leaves the semiotical "intentions" in our political and cultural use of generics out of the picture. These language-born intentions can be discerned from a study of the total logical-semantical network in which generics occur.

Given the formalization$_2$ of elementary logic some hundred years ago, the categories "floating individual" and "free arbitrariness" no longer have any technical jobs to fulfill. However, several minds have since missed their theoretical position in logic as general *forms*$_1$. Both neo-Thomists and German-idealist philosophers have held that an argument from *Some A is B* involves a "middle term" (the exposed term) that stands for an indeterminate or floating individual *"of a general nature*". In order to study such a floating individual Augustinian epistemology is called in. Each such generalized individual should be understood and cherished as a "vague and undetermined subject, taken as a whole" as an "indefinite species" (Maritain), "a logos-bearer" (von Freytag). Our grip of the upper world would suffer from their demise.

By bringing in second-order dialogical rightors and their semantics we have eliminated this technical chimera of pre-dialogical logic — without losing the opportunity of using the traditional vernacular of "generic" terms. Our confidence in our own grasp of the upper world may suffer; on the other hand, the philosophy of

generalized individuals did not support effective discussion, as our second-order rightors do. You cannot have it both ways.

7.2 Intentionality and the subject-object distinction

"Intentionality" is another, epistemological ingredient of the traditional limbo's. In this case we do not (yet) speak of a chimera. However, if theories of intentionality can be pragmatized at all, it will be impossible to do so without attention to the dialogical turn. Dialogical philosophy "deconstructs" the philosophical model of transcendental intentionality, substituting Buber's category of *between* [Theunissen, 1977]. Dialogical logic incorporates this notion: arguments do not stand up by themselves, they are made and judged *between human beings*. The relation between proponent and opponent may be asymmetrical, but, as Lorenz says, the dialogical principle removes or suppresses the division into "subject" and "object" [Lorenz, 1992].

8 "BASIC" AND "PRIVATIVE" CATEGORIES

A logic of formerly privative categories — From privative to primitive

8.1 Formerly privative categories

Like other cultural constructions, the outlook on logic traditionally had a foundation in the *terminus ad quem* — a goal taken from religious aspiration: Belief, Certainty, Truth, Insight, Symmetry, Rest, Agreement, Harmony. A strong transformation consists in shifting from these values as regulating values for logic, to the practical values of contending citizens, lawyers and working scientists: how to reduce initial *doubt* through the elimination of *possible mistake*, *possible fallacy* and *possibly invalid argument* (all of them formerly privative categories).

In addition the strong transformation benefits from a positive attitude towards such categories as Movement and Flux of opinion.

8.2 Old terms: from "privative" to primitive (basic)

Aristotle called a term or category "privative" when — to him — it signalized a mere absence *(privatio)* of one of the categories he and other Greeks philosophers selected as logically fundamental.

Like the Method of Semantic Tableaux, DL breaks through the Greek set of logical fundamentals, putting "logic" on its head: its basic categories often already are chosen from among the *mere privatives* of Greek and later Western logic. In language, privatives are usually indicated by an initial particle *a-*, *de-*, *dis*, *mis-*, *in-*, *non-* or *un-*. If we use capital initials for fundamental (theoretically basic) categories, here are some examples: *dis-Agreement*; *dis-Belief; un-Certainty; im-Precision; ir-Relevance; mis-Understanding; un-Truth; in-Validity*. Much of DL

consists in adopting former “privatives” as one’s basic categories. Their negative names then become a hindrance for the logician, and are better exchanged for non-negative ones (which sometimes have to be constructed).

In the following chart we offer some suggestions. The terms on the left with capital initials are “fundamental” terms of the tradition, those with capital initials on the right stand for concepts that come more and more to the fore as new theoretical primitive terms.[22]

Old Fundamental Concept OLD PRIVATIVE CONCEPT	NEW PRIVATIVE CONCEPT **New Fundamental Concept**
Agreement	dys-conflict
dis-agreement	**Conflict of opinion**
Belief; Self-evident	non-query; in-dubitable
dis–belief	**Doubt**
Certainty	carelessness, unconcern
un-certainty$_1$	**Concern, Caution, Prudence**
un-certainty$_2$	**Heisenberg’s lenience**
Conviction	non-query
lack of conviction	**Query**
Harmony (cognitive),	non-flux, stagnation
dis-harmony	**Flux of opinion**
Precision$_1$	non-ambiguity, . . .
im-precision$_1$	**Ambiguity, Chaos, Fuzziness**
Precision$_2$**, degree of**	in-tolerance, degree of
im-precision$_2$	**Tolerance (of measurement)**
Quietism, Rest, “Logical stability”,	non-Critical attitude
de-nunciation (see right)	**Criticism, Critique, Investigation**
Relevance, degree of	im-perturbance, degree of
ir-relevance	**Perturbation, Dialogical —**
Truth	
un-truth	**Falsity**
mis-take	**Error**
Understanding	mis-upperstanding
mis-understanding	**Blooper, Upperstanding**

[22] A perspective presented to the Henri Poincaré Congress, University of Nancy (1994), published as [Barth, 1996].

Observe that whereas the former fundamentals reflected intended ultimate goals, and hence could be seen as fundamental values, this is no longer the case. The neo-primitives (neo-fundamentals) do not themselves also to stand for positive values (though they may be explained as offering an instrument that demonstrably satisfies positive values or norms).

DL may be called "a logic of former privatives".

IV. Other aspects of logic in dialogue

9 MORE DIALECTICAL RIGHTORS IN LANGUAGE

Pragmatizing sentential logic: connectional rightors — Pluralities of attack and defence — Virtual conjunction — Pragmatizing modal logic: modality rightors

9.1 Pragmatizing sentential logic: connectional rightors

The use of sentence connecting words like "and", "or", "not", ... are understood to carry with them certain possibilities for "reasoning". So-called "natural deduction" consists in the employment of rules of rights and duties for the solitary reasoner. Many deal with them as functions of truth and falsity.

Lorenzen set out to render their meaning-in-critical-use, defining them in terms of the rights and duties they impart on the participants in a critical discussion. We introduced some common "connectional rightors" (as we shall call them) in Section 5.15. Though the natural-deduction rules for these connectors may be called "more pragmatical" than the axiomatic forms of definition that preceded them, the above dialectical definitions are more pragmatical still. At first sight they may not seem to have much to do with the elimination of chimeras, but they contribute crucially to the pragmatical dialogical goals we have already discussed in the sense of being *necessary conditions* for achieving those goals.

9.2 Pluralities of attack

When T_k is a statement of the form $p\&q$ the Opponent likewise has more than one way of directly attacking it:

		$\vdots$	
		$p\&q$	
1	2	1	2
$p?$	$q?$		

In order to defend his/her position as *before* uttering $p\&q$, the Proponent now has to defend his/her position in the first as well as in the second subdiscussion

introduced by O. In other words, in order to invalidate P's position it now suffices for O to refute P's defence in one of the two subdiscussions. So O has a good reason for trying the second possibility of refutation should the first fail.

A similar case arises when the Opponent in the course of the discussion has uttered some sentence of the linguistic form $p \vee q$ and the Proponent at some point chooses to defend his/her position indirectly, by an attack on that sentence.

9.3 *Pluralities of defence*

Now suppose T_k is a statement of the form $p \vee q$. The Opponent has only one possibility of critical attack. Now the plurality of acceptable retort is on the Proponent's side:

$$
\begin{array}{ll|ll}
\multicolumn{2}{c|}{} & \multicolumn{2}{c}{\vdots} \\
\multicolumn{2}{c|}{?} & \multicolumn{2}{c}{p \vee q} \\
\hline
1 & \multicolumn{1}{|l|}{2} & 1 & \multicolumn{1}{|l}{2} \\
 & \multicolumn{1}{|l|}{} & \quad p & \multicolumn{1}{|l}{\quad q}
\end{array}
$$

9.4 *Virtual conjunction (etc.)*

A general standpoint ("thesis" in a wider sense) may contain *virtual conjunctions* (as we may call them) — seemingly separate theses p, q which are not overtly connected into a sentence $p \& q$. It may seem obvious to some, but is not necessarily so to others, that the Opponent can deal with this situation as if p and q were indeed connected by *and*. Similar things could be remarked for other logical constants.

In short, the study of the various kinds of situation that make for the opening of *new sub-discussions* that can be portrayed in *new sub-tableaux* should preceed any attempt to add "new" theoretical topics.

9.5 *Modality rightors*

In two stages the metaphysical semantics of "possibility" and "necessity" was bent away from the impractical category "essence", first, by a semantics based on the notion of a multitude of "possible worlds". This extricated the modal categories from their earlier connection with classification and moved them towards a wider modern logic. From there they could be reinterpreted, without theoretical loss, as pertaining to a range of "discussion levels" ([Inhetveen, 1982], with reference to Marcinko; [Krabbe, 1986]).

This pragmatization of modal logic ties up with the following distinction of "dogma" and "conviction": "a dogma" may be defined as "a conviction that one is not prepared to discuss at all". The only dialogically befitting dogma then is: *"There are convictions of mine — though I do not know which ones — in which I may be wrong"*.

In this spirit some logicians have exchanged the semantical-logical term "universe of discourse" for "universe of discussion".

Deontic terms such as "ought", "obliged", "duty", with "may", "free" and "right", "permitted" and "forbidden", can undoubtedly be treated in similar manners, resulting in *deontic rightors* (my term).[23]

Paraconsistency: Mackenzie and Priest [1990], most recently by Rahman and Carnielli [1998], apply a dialogical perspective, the latter authors in DL terms.

10 ADDITIONS TO METALOGIC?

The Theory of Games — Olga and the Pope — Which system of modern logic is dialogically more natural than the other ones? — To which extent are dialogical, semantical, deductive (etc.) conceptions equivalent?

10.1 The theory of games

In 1958 Beth gave a lecture to mathematicians in which he said:

> "[We must] try to eliminate such subjective, and sometimes even mystical, elements as can be found in most intuitionistic writings. ... [An] explanation in terms of the theory of games seems to be the more convenient in various respects. We imagine the following game between two players, A and B ..." [Beth, 1959b].

That same year Lorenzen wrote:

> "It seems to me that the understanding of logic that I develop below ... is independent of the peculiarities of the philosophizing of Brouwer about thinking and speaking. ... We imagine 2 persons ...".[24]

10.2 Olga and the Pope[25]

Perhaps the very notion of "metalogic" could do with a few pragmatical additions. As an illustration of the importance of being able to formulate logical rules the dialogical way, it may suffice to imagine a discussion between the Pope (or other person with the power of influencing the world by decree) in the role of Proponent. Now suppose Olga, or Otto, would like to act in the role of Opponent:

---

[23]Cf. also [Lorenzen, 1969].

[24]Beth [1959b]; see Velthuys–Bechtold [1995, No. 543]; Lorenzen around 1958. In the nineteen-seventies Hintikka reconstructed Beth's semantical-tableau method in a manner that introduces one role, "Me", in which someone operates versus "Nature" (to call this a role would seem forced). His semantical material was not yet that of human dialogue [Hintikka, 1979, n. 11], nor did he at that time pay noticeable attention to the wider potential of logical revision the earlier contributors — Beth, Heyting, Lorenzen, Naess — were looking for. This reconstruction is a concrete case of no-role-to-one-role pragmatization ("weak pragmatization", above, Section 3). Such a transformation of objectivist theory into a Kant-like epistemic theory of logic could also be called a *Kant transformation*.

[25]Cf. Barth and Krabbe [1982].

Olga	Pope
Con	T_0
$\mathrm{crit}_1(\mathrm{T}_0)$	

Will the Pope react? Does the linguistic form of the Pope's thesis allow for a critical attack at all? Or did he choose a way of expressing himself such that nobody would be able to say how to open a critical attack on *any* thesis of such a linguistic form?

10.3 Which system of modern logic is dialogically more natural than the other ones?

Lorenzen's university teacher, operationalist philosopher of science Hugo Dingler, had looked for first principles in semantics. Lorenzen himself was inclined to seek the firmest ground somewhere within mathematical thought, of which several creeds and versions are known. Heyting and Beth had given Brouwer's creationist *anti-dialogical* intuitionism more dialogue-prone interpretations, so that you could now defend a *dialogical constructivist* outlook. Lorenzen's operationalist training gave him the idea that, from a dialogical point of view, in its two-valued formal$_2$ logic was somehow "less obvious" than formal$_2$ logic in its weaker constructive form.

However, already from the beginning (around 1958) Lorenzen realized that the whole difference between the two logics boils down to the question of one right more or less to the Proponent [Lorenzen and Lorenz, 1978, 8]. This right — to make use, at any stage, of everything the Opponent said in the discussion so far — makes the discussion (as it were) "two-valued". In general, the rule allows the Proponent of the initial thesis *to return from the defence of a later subthesis to the defence of an earlier one*. This rule can be shown to correspond to the assumption of objective "truth-values" for the statements involved.[26]

In other words, the "weaker" constructive logic is, in dialogical terms, *stricter* than the two-valued one from the point of view of the Proponent. Hence in a constructivist dialogue the Opponent sometimes — notably where notions of non-denumerable infinite sets are involved, a problem sphere we need not worry about here — has a better chance of not losing the argument.

With this additional rule the Proponent can assume p whenever the Opponent rejects *not* p.

Constructive dialogue logic does not deal with the truth-falsity distinction. This does not mean that it knows no definition of negation. Its notion of negation consists in one assumption — one conventional rule, namely: when an *absurdity* is stated, the critical listener of the moment, whether Proponent or Opponent, may without loss subsequently utter any sentence whatsoever ("if one absurdity is allowed, then anything goes").

[26] It also mirrors a rule already known from deductive tableaux, cf. Beth's lecture in 1958[1959b].

Stricter yet (to the Proponent) is Johansson's "Minimal logic". Minimal logic is defined as what remains when this possible convention is thrown out; instead one agrees on another rule: "As soon as someone utters an absurdity the discussion is closed, with loss for the other party".

Anything weaker (stricter to the Proponent) than that, Johansson says, is not worthy of the name "logic". Hence the name "Minimal logic".

Minimal dialogue logic is, therefore, a slightly simpler system — it contains one rule less — than constructive dialogue logic, which is slightly simpler than two-valued logic. It does not follow that one of them is "more natural". In practice, choice will depend on agreement (Section 5, on conventions).

For the purpose of practical application the following is of great importance:

It is easy to show that negation in one of the traditional "logics", viz. the Hegelian logic (also called his dialectic), is even weaker than Minimal logic, and that it does not allow for rules for critical discussion.

10.4 To which extent are dialogical, semantical, deductive (etc.) conceptions "equivalent"?

If we understand the "equivalence" of methods as the algorithmic convertibility of an application of one method into an application of another method, then the answer is yes. The condition is that one refers to well-defined and well-known formal$_2$ analytical methods and definitions. The question is often put in terms of the "completeness" of one method relative to another (historically, most often of deductive methods relative to semantical and model-theoretical methods). One can speak of this concept as the concept of *algorithmic formal$_2$ equivalence*.

The algorithmic equivalence of the dialogical-tableau method for first-order DL* with the semantic-tableau method was clear to Lorenzen from the outset, as he explains in a letter to Beth in 1959. The latter method is demonstrably formally$_2$ equivalent (aequipollent) with the method of deductive tableaux, the method of natural deduction and the earlier method of axiomatic deduction [Beth, 1962].

Taken together, this is to say that a desired conclusion can be shown to follow deductively from given premises if, and only if, a Proponent of the said conclusion *has a winning strategy vis-à-vis any Opponent who concedes these premises, in a discussion that has this desired conclusion as its initial thesis*. Formally$_2$, it is therefore fully warranted to allot the central position in logic to the dialogical point of view.

However, "equivalence" of methods can also be understood as the much wider concept that fully takes in the practical perspectives we have dealt with above. With this concept of equivalence in mind in our theory of logic we must conclude that the deductive methods — the axiomatic procedures, the methods of natural deduction — are certainly not theoretically equivalent to the dialogical procedures. In dialogical logic, formally$_2$ nothing is lost; pragmatically it opens new vistas.

In the service of clarification through "precization" of separate principles of a

moral or of a political character, an axiomatization of systems of such principles in a given discourse is indispensable. Current examples of elementary model theory, usually taken from mathematics, are in need of complementation with examples from systems of scientific, moral and political principles.

11 SUITABLE CASES FOR TREATMENT

Extending the accessibility of logical theory — Great expectations — Missing: a survey of practice in various fields — Logical self defence, political discourse, and education — Dialectical forecasting

11.1 Extending the practical accessibility of logic and logical semantics

Of many branches it is to be expected that their practical potential will be be better understood after a pragmatization in our present sense. Among the more promising candidates let us mention: ambiguity and vagueness; definition; deontic logic; formal_2 logic of higher orders; game-theoretical semantics; game theory; inductive logic and reasoning under uncertainty; interpretational and conversational principles; intertextuality; metalogical parts; non-monotonic logic;[27] preciseness; pro-et-contra, pro-aut-contra analysis; rationality; refutational logic; relevance; truth concepts. To carry this out is professional work. As we have seen, in some of these fields the work has begun.

A particularly consequential field in this connection is that of the *logic of constructing counterexamples* (Beth), developed about the same time as Łukasiewicz' *refutational logic* and related to it in spirit. Whereas in formal_2 logic the dialogical representation mirrors Beth's refutational method of semantic tableaux, strangely enough Łukasiewicz' approach has not yet been approached as an intersubjective problem.[28]

The logical technique of *reductio ad absurdum*, beginning from a false assumption, has been the object of criticism by Schopenhauer, Goblot and others, who seem to have experienced this mode of "reasoning" as immoral.[29] A dialogical pragmatization vindicates it.

As to *fallacies*, one should take to heart Woods' reminder that the concept of fallacy is "not a unitary concept", together with his corollary that no unitary treatment of types of fallacy should be expected [Woods, 1994].

Since 1974 at least, Woods, in part in co-authorship with Walton, offers acute practical analyses appealing to various branches of modern logic. It seems to me that an effective pragmatical understanding of what is involved in the various fal-

[27]Cf. the definition of cumulative logic in [Barth and Krabbe, 1978].
[28]Cf. [Tamminga, 1994], which may be read as a preparation.
[29]This is discussed in Ch. III of [Beth, 1970].

lacies can only be had *via* a dialogical pragmatization of the logics concerned (inductive logic, relatedness logic, etc.).[30]

There is a lot to be done. That Montague's *pragmatization of logical semantics* can be yet further pragmatized — in the present sense — is demonstrated in [Barth and Wiche, 1986].

Other fields argumentation and rhetoric; cognitive science and theories of "representation", linguistic and mental; education and didactic authorship;[31] ethics and morality; legal theory and philosophy of law; political science, propaganda and political culture, and prejudice; psychiatry; sociolinguistics, with theories of inclusion and exclusion (dialectical fields, verbal self defence); theology, particularly theological discourse among humans.

Toulmin's notion of the "field-dependency" [Toulmin, 1958] of argument may be put in dialogical terms; see, for instance, [Johnson, 1993]. The same goes for applied and other applicable parts of mathematics.

11.2 *Great expectations*

For each new addition to logic or to logical semantics, some persons of a metaphysical bent seem to expect it to work wonders. At the advent of many-valued logic the hopes were enormous. Modal logic was expected by some to be able to improve medieval efforts to demonstrate the existence of God. More recently fuzzy logic was supposed to be revolutionary. Above we mentioned a case in which mathematical category theory was used in the hope of reverting to pre-modern representational forms bolstered by the high-status field of mathematics.

Let us repeat that dialogue logic is not recommended here as a panacea for all problems. It had better be treated as an ingredient giving form to Albert's "methodological revisionism" programme [Albert, 1996][32] — a promising remedy against relativism and total skepticism alike.

Extension of this empirical-revisionist outlook to logic itself was not originally part of the critical-rationalist programme. In the late nineteen-thirties the young Naess came close to it, but Beth and Lorenzen seem to have been the first to introduce it more explicitly as a criticism of the philosophy of logic itself.[33]

[30] As Martens and I suggested, the discussion of fallacies in[Johnstone, 1959, 3f.], though admiringly concerned with what we call the Opponent's position, is still unclear, because Johnstone does not sharply distinguish between *the roles of opponent and proponent in respect of a certain thesis* T. Cf. [Barth and Krabbe, 1978, 89f.].

[31] Compare van Bendegem [1982; 1993], Barth [1977].

[32] In English: "The myth of the framework and modern anti-realism. Criticism of the idealist relapse in contemporary thought".

[33] Cf. Naess [1956; 1992a; 1992b].

11.3 Missing from the study of logic: a survey of logical practice in various fields

Whenever the logic of science concurs with the spirit of dialogue, conversational principles, as Grice calls them, can be taught from scientific practice. Debate may seem a substantial component of all scientific discourse, but not every philosophy of science and not every practicing scientist subscribes to this outlook.

A good example of a "conversational principle" in scientific logic is the practice of *systematically referring to related scientific material* — to finding places; to origins, theoretical and terminological, whenever the theory is not yet internationally established; to opponents and future opponents. *The practical logic of mathematical philosophy* shows enormous differences here, even within one and the same country. For exemplary cases of a dialogical horizon, notice the enormous bibliography of works discussed in [Beth, 1959a], as well as the civil dialogues between imagined mathematical colleagues of differing persuasions, one of whom introduces constructivist mathematics — not yet itself dialogically formulated — to the other ones, in [Heyting, 1956] (the title is dramatically misleading). The peremptory intuitionist Brouwer, on the other hand, "seems to imply that only those with whom he 'can do battle' are worth investigating" [Stigt, 1990, p. 129]; cf. Lorenzen in [Lorenzen and Lorenz, 1978, p. 2].

Biological, political, and geometrical discourses have, unexpectedly perhaps to some, a logical history of ideas in common. In all of them: look out for "floating individuals", "genera", "two-ness", and "types". Political discourse is studded with uses of the generic sentence form: *the (general) A*, *the (general) A as such*, and so on. It has been detrimental to an understanding of the importance of this fact that most frequent example given has been "the Triangle", for this is the one case in which a premise *every A is B* and a premise *the (general) A is B* are presumed by everyone to allow for the same conclusions. On the other hand, neither in practice nor in theory has the sentence *The Rose is beautiful* been regarded as equivalent to *All roses are beautiful*, nor has *The Jew is so-and-so*, used by anti-semites, been taken by them or by anyone else as equivalent to *All Jews are so-and-so*, hence refutable by means of a counter-instance.

As for biological discourse two remarks are due. The first is that today the concept of genus and species is much criticized even among theoretical biologists. The second, that its dangerous uses is most visible within sociobiology, where it produces obstacles for effective discussion (cf. Kitcher, Van Laar).

As for political discourse, there is a remarkably high frequency of generic uses of terms particularly in extremist politics — fascism, nazism, communism, and other varieties.

In physics, generic categories have given way almost completely to the category *Function*.

The theory of special relativity is an interesting case. The theory itself may be in no need of dialogical reformulation, but the understandability for non-physicists might well be enhanced by a reformulation in the terminology of dialogical logic.

Einstein's *relational* theories of physical relativity exemplify the value of theory pragmatization as a means to combat "relativism" as that word is used in the contemporary discussion. They demonstrate that by relating descriptions of measurement to the measurers' co-ordinate systems and their relative velocities it becomes possible to maintain a notion of objective truth in the sense of intersubjective truth. They go against the present-day fashion of rejecting "objectivity" in favour of *laissez-faire* relativism and subjectivism that invite overall attitudes (cf. Albert's 1997 attack on the "myth of the framework" in modern anti-realism).

In atomic physics an attempt has been made to clarify the theoretical particle-wave antagonism by representing the experimental situations in Lorenzen tableaux [Mittelstaedt, 1978].

"Reasoning". One author looked for a connection between "logic" and reasoning without finding one [Harman, 1984]. Mackenzie concludes [1989] that one has to take in also the logical constants — which is precisely what Lenk did in 1968. Here the neglect of major non-English language publications is particularly grating.

Mathematics. Though no one has as yet statistically proved it, interest in the logic of dialogue seems to be positively correlated with an interest in systematical empirical research concerning logics as used in practice (as in [Barth *et al.*, 1992]). Naess [1953] is a case in point.

"What happens when X sets out a mathematical proof to X? How can Y put forward questions to X in such a setting?" Van Bendegem asks. Starting out (in 1982) from the correspondence between Frege and Hilbert (on the foundations of geometry), via an interlude (1985) where he attempted to see the dialogue as a filling in of missing steps in a "top-down" analysis, he returned to the empirical approach, collecting elements for a realistic model of mathematical practice (1985, 1993). Kitcher has subjected both mathematical and sociobiological discussion to a careful analysis.

11.4 Logical self-defence, political discourse, and education

A contemporary textbook on informal logic carries the significant title *Logical Self-Defence* [Johnson and Blair, 1994], most recent edition 1994. This title contrasts interestingly with Plato's and Augustine's epistemology of intuition/insight. The necessity of verbal self-defence invites a scrutiny of the "dialectical fields" of attraction and repulsion, or, inclusion in and exclusion from *participation* in dialogue.[34]

11.5 Dialectical forecasting

In politics, political science and sociology the notions of "dialectics" are often traditional, non-verbal, and anti-dialogical. The same holds of political and theo-

[34]Barth [1977; 1985d]. Prejudice is discussed in [Barth, 1981; Rose, 1990].

logical practical discourse. Self defence by logical means is not contemplated.

In order to make dialogical logic ready for political use it will be necessary to familiarize the population at large with the unmistakable dangers of generic argumentation. Schools must explicitly recommend a switch to other vernaculars when available — for instance, a switch to expressing opinions in terms of functions and vectors, etc. In addition they should initiate their students into the differences between the many ranges given to the hidden second-order variable "M" in practice (see Section 7.1). Then one can go on to recommend that listeners in a debate ask for a specification of this variable; demand a clarification as to whether *every* X or only *most* Xs or ... must be Y, or perhaps "can" be Y, or "is potentially Y", ... And: *don't forget to make use of the possibility of casting doubts on M itself*!

Some critics do that already, in very involved academic discussions, but they are seldom successful. Many, many more do not even know how to start such theoretical discussions and go through life in an ocean of generic locutions with monstrous consequences, unaware of what is, or ought to be, their dialogical rights. Theological discourse among humans comes in here, too.

Before the gargantuan bloodlettings in the 20th century, logicians were not only without educational influence, they were totally unequipped to make use of such influence had they possessed it. Yet there is no reason at all why forecasting of oncoming semantical/dialectical storms and the direction they come from ("dialectical forecasting") should not be just as feasible as forecasting of ordinary weather.[35]

V. Cross-cultural logic

12 LOGICS IN INTER-CULTURAL DEBATE

"Between the cultures there can be no reason" — Logic as the critique of logics. Empirical methods — Imaginal functional compensation — Logemes

12.1 "Between the cultures there can be no reason"

As criticist thinker Hans Albert recently pointed out, this dictum of Oswald Spengler's was a main source of the relativism, pessimism, and hopelessness that counteracted effective intervention in the development in the interbellum. Albert emphasizes that today a similar philosophy of cultural difference influences our outlook on extremist beliefs.

Professional logicians aiming at the applicability of logic in political practice will sooner or later have to address the fact of (il)logical variety. There is no need to claim at the outset that all logics have, or that they do not have, a common nucleus, or that all extant forms of mental or verbal logic are of equal value. In

[35]An image that is used both in [Barth, 1982] and in [Woods, 1993].

the opinion of the present writer, some are remarkably better than others. But variety there is. An empirical interest in this variety may take the form of research into the history of logic that does not exclude dubious developments (non-whig history of logic, cf. Russell, Beth, Geach, Rescher, Hacking, Kitcher and others). Thanks to Foucault's and Rorty's writings the importance of historical studies for the pragmatically minded has become clearer to a large audience.

12.2 Logic as the critique of logics. Empirical methods

In sum: in the interest of the further development of a dialogical logic knowledge of different logical apparatuses (and of their respective powers) is required. Notice that to have an *empiricist attitude* towards the foundations of logic does not necessarily mean that one takes an interest in the human variety in matters of logic and the application of *empirical methods* for the study of this variety. Logic as the critique of extant or historical logical theory will have to develop from empirical descriptions.[36]

Here the study of Arabic, Chinese and Indian texts on logic comes in. In this respect probably no single person has done more real work than Nicholas Rescher, who significantly has also discussed dialogical procedure in epistemology [Rescher, 1977]. Furthermore, field work reports from cultural anthropologists delivers material and important insights — Malinowsky, who inspired Naess; Durkheim; N. Sombart. Sometimes these insights are strikingly similar to studies by working logicians:

12.3 Imaginal functional compensation

Consider two different systems of linguistic (or merely cognitive) representation, both intended to assist us in our attempts at producing good arguments; let us call them S1 and S2, and let "f" be a variable for, as Geach puts it, the "logical jobs" (functions) for which some linguistic representation or other is taken to be needed. (Examples of such logical jobs abound: the task of expressing non-adherence to a statement; the task of expressing a belief in some kind of generality behind a specific utterance; and so on). Now assume that in S2 a representational category or principle R is employed in order to do a certain logical job whereas in S1 this R is unknown. For an adherent to S2 this may seem to indicate that adherents of S1 have no linguistic and no cognitive instruments meant to carry out that logical job — the (sub)culture in question may not even have discovered that in the process of distinguishing between bad and good arguments, this is a logical job of importance.

The presumption about S1 may, but need not, be correct. The (sub)culture to which S1 belongs may have hit upon another way, perhaps hopelessly clumsy, perhaps very smart, for taking care of f. It may possess some principle or category,

[36] This use of "logics" in the plural has not been common. It was assumed by the present author in [1972; 1975], by Haack in [1974], but was probably occasionally also used earlier.

say $R*$, that *in intention* — from the point of view of at least some minds in that (sub)culture — does the job for them, and so far does it to their content. Briefly:

> In such cases adherents to S2 ought to realize that R$*$ offers adherents to S1 some amount of *imaginal functional compensation* for the absence of the category or principle R as employed in S2.

This principle may be discerned in the writings of the sociologist Durkheim. We might call this the Durkheim–Beth principle.

S1-people may well regard R, when they hear about it, as superfluous. They may also find S2 sadly inadequate in as much as R$*$ or parts of it that are also used for other purposes, are entirely missing. One example: pre-Fregean and pre-Russellian logic (S1) operates with a monadic syntax and represents non-symmetrical relations hierarchically (R). Later logic (S2) represents symmetrical as well as non-symmetrical relations through polyadic predication (R$*$), and has no need to invoke hierarchical order for that purpose.

12.4 Logemes

The totality of concepts and conceptions that carries a logic in practice may be called a *logeme*.[37] Without first-rate knowledge of the logemes involved, the endeavours of dialogue logicians working towards an intercultural or inter-subcultural dialogical logic are doomed to be abortive.

13 DIALOGUE LOGIC IN PRACTICAL ETHICS

Criticism and Holistic You-ness — Accountability (4) — "Let us transcend our subjectivity!" — Understanding the urgency of logeme shifts, a dialogical requirement — Rules-of-the-thumb for forging a logeme shift in the Western hemisphere

13.1 Criticism and Holistic You-ness

Not all talk of "You and I" is meant to open the doors for dialogue and due process. Both theoretical and practical ethics has suffered from a holistic concept "You" that goes contrary to dialogue-logical categories (you-as-a-critic). In that ethics, "the living You of *each* fellow countryman floats into my own soul and grows there to a larger, holy You that shines above all of us in unison," a Nazi philosopher wrote in 1936; to confront this 'holy You-ness' with individualistic ideas of a critical or rationalistic nature would amount to blasphemy [Kiesewetter, 1996, p. 258].

[37] Cf. my [1985c; 1993; 1996].

13.2 Accountability (4) Enclose your opponents in thematic discussion

In politics and ethics, substitute dialogue logic for the ancient epistemology of guilt-causes. The dialogical theory is the only theory of logic that incorporates this norm in its very machinery.

In an anti-dogmatic setting, an emphasis on the true-false distinction goes some way in the same direction.

Logemes that contain at least two dialectical (dialogical) roles allow for you-as-my-critic and for the practice of *verbal negotiation*, where attack and defence pertain to entities in the verbal space between the speakers, and only to these. Let us refer to this set of categories as B. Facing new logemes based on these categories one finds logemes that give pride of place to the much older category of "guilt-causes" [Benthem van den Bergh, 1980], which we may call A. As a rule, the latter also assume one final goal, that of "the" happiness or salvation of a certain holistic entity, and endow this entity with a superindividual ”You-ness”. Let us compare, one by one, the various features of the struggle that can ensue between actors who depend on some A-related logeme with those that are open to actors who are trained and experienced in the use of B-related logemes:

(1A) "I accuse you (or You) of being morally guilty/I indict you/I dispute your conditions of existence." Contrast:

(1B) "I accept you as a discussion partner in a critical discussion according to rules to be agreed upon in advance, where I shall take on the role of opponent to your utterance U of your unfortunate opinion (which, having been uttered, is present outside you, in addition perhaps to being still inside you). My first critical remark concerning your utterance U is the following: ...... "

(2A) "I have brought a guilt charge against you, I have fought/eliminated you/ your conditions of existence." Contrast:

(2B) "I have successfully countered your defence against my criticism of your utterance U, formulated in this particular discussion or belonging to your concessions. In other words, I have won this discussion with you."

(3A) "I shall defend myself as best I can against the effects and consequences of this guilt charge, or: against this attack on my person/my conditions of existence." Contrast:

(3B) "I accept your challenge, i.e., I accept you as opponent of my utterance U, hence as reviewer of its tenability against verbal criticism."

(4A) "I confess guilt." Contrast:

(4B) "In the light of the debate we just had I accept your criticism of U" (or perhaps: "I immediately accept . . .)."

(5A) "I have lost in the struggle against my person/my conditions of existence." (Or: silence; the dead do not speak.) Contrast:

(5B) "Today I have had to give up my defence of my utterance U, that is to say: I have lost the discussion we had about my opinions (as I expressed them in U). I admit defeat in this discussion about U with you personally."[38]

13.3 "Let us transcend our subjectivity!"

Some persons react to the notions of verbal "attack" (challenge) and "defence" as if there were no essential difference between an attack in a critical dialogue and the use of violence. This attitude towards B-related frameworks encourages and facilitates the A-related activities enormously, in practice as well as conceptually. A notion of "cooperation", treasured in the Christian churches since St. Augustine, is sometimes recommended as an alternative to the philosophy of verbal struggle. Notice that the notion of "cooperation" was generally employed in extremely militaristic proto-fascist, fascist and Nazi philosophies, as well as in communist and other totalitarian regimes, as the positive alternative to debate, which was outlawed as a sign of illoyalty.

"Let us transcend our subjectivity!" Lorenzen writes in [1969]. What we have called "externalization of dialectics" consists in the shift from (A) to (B).

13.4 Understanding the urgency of logeme shifts, a dialogical requirement

For cross-cultural, cross-subcultural utility there is a need for new and clear conventions — global, but acceptable to all parties implied; regulated, but adjustable.

It seems inescapable that, in the interest of the construction of globally effective dialogue logics, the various parts of the world weigh their own adopted modes of argumentation against the requirements of effective dialogue with "aliens".

Not all logemes can and will support a system for critical dialogue. In the course of the last century Europe, for instance, has been flooded with counter-dialogical mentalities. Let us therefore sum up some rules of thumb for how to thoroughly extricate the Western science of logic from Western versions of Wittgenstein's fly-bottle: the counter-dialogical logemes in which logic has been ensnared after the "decay of dialogue" around 1300 [Ong, 1983]. Familiarity with logeme disparities may open the eyes of logicians for the desirability of logeme shifts, at home as well as abroad.

[38] This exposition is inspired by conversations with G. van Benthem van den Bergh; cf. his [Benthem van den Bergh, 1980].

13.5 Some rules-of-the-thumb for forging a logeme shift in the Western hemisphere:

(1) *Concerning the traditional Goal, and "privative" categories.* In matters of logic, shake off all remnants of the philosophical fascination with the *terminus ad quem* — "the" locus of Rest, abeyance; and focus instead on the *termini a quibus*. Give analytical pride of place to categories that have been treated as "privative", in this case: Movement, redefining Rest as an absence (non-Movement).

(2) *Concerning "Being" and the linguistic turn.* Exchange the project of a logic of Being for that of a logic of Statements. As basic (initial) concepts, exchange the notions of Harmony and Logical stability for the notions of Controversy and Flux of opinion.

(3) *Concerning interiorism and parliamentary activity.* On the backdrop of the "logic of Being" monologue and interiorism got a chance, and took it. So remember to eliminate logical remnants, possibly subconscious, of fundamental interiorism (St. Augustine; Luther), of ideologies of "the I", of "bearers of Logos" and "Silence" and "Creative Knowledge" (Fichte, later German idealism including our century). — Also eliminate remnants of the gnostic-Neoplatonic category Opposition/Battle in Being, or in the Course of the World, as in Manichaeism, gnosticism, and Taoism. This is the category "Contradiction" (German: "Gegensatz") in "Being", exemplified by Hegel, Lenin, Quisling and Mao. Replace with the notion of *Opposition of roles in verbal intercourse*.

(4) *Concerning pragmatization.* Here are, in alphabetical order, some productive categories central to the general undertaking of pragmatizing extant scientific theories, though alien to pre-dialogical logic: *Conventions* (transactions); *Coordinate systems*; *Functions* (for early emphases on the importance of this category (cf. [Cassirer, 1910]; *Problems* (emphasized in [Barth and Wiche, 1986]; *Relations* (not only binary, and not only hierarchical); *Roles*, as in semantic roles (speaker, listener) and parliamentary roles (Proponent, Opponent).

(5) *Concerning non-symmetrical categories.* As for binary categories, exchange the emphasis on symmetrical categories like Similarity, Difference, and Exclusive disjunction, for attention to non-symmetrical and asymmetrical categories, e.g. Comparative relation, Ordered pair, Not both, At least one (or At most one, in combination with unary Not). Replace an epistemology of individual beings as symbolizing "the Universal" or "the Typical" by virtue of their "similarity" to the latter, by a language and a logic with separate symbols for Individuals and rules that employ such symbols.

(6) *Concerning fundamentals.* Abandon the dogma that harmony, precision, relevance and God-given validity are our natural logical fundamentals. Theoretical pride of place should be given to certain categories that traditionally have been belittled as privative: *Controversy, Falsity, Risky reasoning* and attempts at *Falsification*, with *Counter-example, Counter-argument, Counter-model, Dis-Proof, Poly-interpretability and Ambiguity* (Naess); *Dialogical perturbance, Agreed-by-all Annoyance*, or *im-pertinence*, etc.[39]

14 CRITICISMS OF DIALOGICAL PRAGMATIZATIONS

Overt derision — Model-theoretical Limbo — Dialogical indifference

14.1 Overt derision of dialogical endeavours

By demonstrating their eliminability we have shown the State-A relatedness of certain concepts or categories: arbitrary individual, infinitesimal (magnitudes, later: numbers).

Twentieth-century examples of derision of discussion abound. We have mentioned the overtly anti-dialogical ambition in Brouwer's creationist intuitionistic philosophy of mathematics [Stigt, 1990]. A political example from the same period is Count Hermann Keyserling's well-known *School of Wisdom* in Darmstadt, Germany, where, for the sake of Wisdom, discussion was entirely prohibited.

15 MODEL-THEORETICAL LIMBO

Two writers on mathematics strongly *prefer* Robinson's model-theoretic definition of neo-infinitesimals to Weierstrass' dialogue-prone renewal of the Calculus as little more than an irritant, calling it a "tongue-twister" which is accepted only "for the sake of consistency" [Davis and Hersh, 1972, 245f]. Robinson's demonstration of the possibility of defining "infinitesimal" in such a way that their use does not entail inconsistencies is elegant; however his neo-infinitesimals are also acknowledged to be superfluous, i.e., neo-chimeras.

Some years ago a philosopher propagated a return to "arbitrary individuals" in logical semantics [Fine, 1985]. As in the former case, a "semantical model" for (human) logic is introduced from nowhere, featuring neo-arbitrary individuals as the referents of instantiated variables. No doubt it is technically possible to represent modern logic in an ontologistic manner on a two-dimensional "mental screen". It is fully possible to teach logical instantiation in terms of artefacts called "universals" or "arbitrary individuals"; however, this seems to require that the learner is, for the time being, uninterested in logic for dialogical use.

---

[39] Naess: *usaklighet*, German: *Unsachlichkeit*.

Here is an analogy. Suppose no physicist had been interested in ether *winds.* It would then in all likelihood be feasible to postulate a model for electromagnetic theory containing an artefact with the name *ether*, which could be given — by definition or postulate — some of the properties the old "ether" was supposed to have. However, the question of whether we really need such a category would still be a reasonable one: Can physics be done without a neo-ether? Can logic be done safely and simply without neo-infinitesimals and neo-arbitraries?

15.1 Dialogical indifference

If so, are we then not entitled — thinking of neo-arbitraries — to speak of a certain philosophical drift, now on the wings of model-theory, towards a neo-ontologistic (rather than intersubjective) form of representation, and to query into the reasons for this drive?

Indifference to dialogical possibilities has also made possible a recent plea for a return to generic terminology, now to be based on the powers of a modern branch of mathematics called category theory [Ellerman, 1988]. The attempt was later shown to be backed by a mathematical mistake [Paalman de Miranda, 1990]. We see, then, that — on no empirical basis whatever, with nothing but disdain for human practice — sheer mathematics has on occasion been used as de-precization of the actual problem in question.

A counterattack is needed. As a point of departure, take the criticism of Augustinian fundamentalism given by existentialist philosopher Karl Jaspers. In Augustine's philosophy verbal intercourse is subjected to the conditions of authority, always able to break off communication in favour of violence [Jaspers, 1967, pp. 164, 171].

The critics of dialogue-prone endeavour must be inspired by other ideals and norms. One such norm that may be involved here is an esthetical one: find the *shortest* representational form for the mathematical theory of "proof". The dialogical representational form may not be the shortest — but what is a proof? "A proof only becomes a proof after the social act of 'accepting it as a proof'", Manin wrote in [1977].

16 NORMS AND THEIR STANDARDIZATION

A normative reconstruction of a possible set of rules for DL — Principles of "Effective dialectics" — Expediting the enterprise — Cooperation — Real conventions, or the enhancement of logical understanding through international standardization

16.1 A normative reconstruction of a possible set of rules for DL

The rules of DL as well as the formal$_2$ rules of DL* will probably be more easily accepted if they can be seen as a solution of the problem of *implementing* norms that people are likely to endorse.

When norms are made explicit it should become easier to achieve agreement about rules that implement them, at least in the long run. If not, familiarity with the method of relating rules to norms will at least illuminate the situation. This may encourage the societies involved to collectively undertake a revision of norms for inter(sub)cultural logic, rather than rely on principles of religious or political authority, intuition, or self-evidence.

Some of the rules have been discussed above. They are seen to satisfy a set of norms that comprises the following (we do not here mention all):

- The dialectics should be a dynamic feature of culture. The system of basic rules as a whole shall be designed so as to *allow and perhaps promote revision and flux of opinions*.
- The dialectics should be systematical. (a) The Proponent of a challenged initial or intermediary thesis shall be given the opportunity to defend/justify it by making yet *another statement*, provided the Proponent will assume the same responsibilities ("pro-position") vis-à-vis the new statement as towards the previous one. (b) This opportunity shall be *realistic:* in some cases it must be possible for an attacked statement to be *defended/justified unconditionally*;
- A system of dialectics should be thoroughgoing. At every point in the discussion the Opponent (of the initial thesis) shall have the opportunity to test the tenability-against-criticism of the Proponent's last statement so far; and the Proponent shall have the opportunity to attempt a defence/justification of that statement in all possible manners;
- A system of dialectics should be orderly. At the completion of any stage in any chain of arguments, the positions and obligations (and thus the rights) of each party should be determinable from the discussion so far (including this stage); and this should also hold for the relative urgencies of the parties' various obligations, if there be more than one;
- Accountability of the loser: If, in a certain chain of arguments, one party has lost or exhausted its rights within in that chain, then this party shall — publicly, if the other party so wishes — express that the other party has *won* this part (this chain of arguments) of the debate; (etc.).

For example: given the norm that our dialectics should allow for a flux of opinions, we shall want to reduce repetitiousness to a minimum, so we need rules that rule out unnecessary verbal repetitions. On the other hand: our norm of thoroughness is well served if we let discussions split up into chains of arguments. These

may be initiated by the plurality of critical moves that are possible with respect to the other party's uses of *all* or of *and*, or by the plurality of moves in defence of one's own uses of *some* or of (veljunctive) *or.* This means that in a given situation the rules should permit each player to try out every possible criticism once (opinions are divided as to whether and when we ought to add: "but only once").

16.2 Interpretational ("conversational") principles of an effective dialectics

In our rules-of-the thumb above the notions precision, relevance and validity are not discarded; they are reborn as mere absences (neo-privatives).

On "effective discussion" a number of principles have been formulated [Naess, 1981; Gullvåg and Naess, 1996]. It is a disgrace that no detailed comparison has been undertaken between Naess' early ideas on logic and dialogue (in English 1966, 1992) and for instance Grice's. This fact, if anything, testifies to the need for more depth and new academic norms in practical logic.

As synonyms for "effective" Naess uses "serious", "non-tendentious", as well as "relevant", but writes:

"We shall consider *irrelevance* under six main headings".

In short, it is *irrelevance* that has pride of place. The definition he adopts for "relevance" is that "a matter X is relevant" in a given discussion turns out to mean the same as "X infringes none of the following principles for *irrelevant discussion*". Relevance is a privative term.

Naess' principles for relevant discussion turn out to concern *irrelevant* verbal behaviour. He lists the following rules of avoidance of such behaviour:

(1) Avoid *tendentious* use of contexts,

(2) avoid *tendentious* renderings of other people's views,

(3) avoid *tendentious* ambiguity,

(4) avoid *tendentious* first-hand reports;

(5) avoid *tendentious* argument from alleged implication.

Instead of "relevant" we find "tendentious" as the basic term. The drift is clear: the participants' use of contexts, their renderings of each others' and other people's views, their own first-hand reports of events, and their reference to implications from known material should *not be such that someone might call them tendentious*. "Tendentious" and "irrelevant" are not synonymous. "Relevant" is often taken to refer to semantical proximity, leading to never-ending impractical linguistic speculations, whereas "tendentious" is closer to "irritating" and "rude". Now what is to count as irritating or rude is something that may, to a large extent, be decided upon in advance. The decision will not be one for all times — not a matter of objective Language, but rather a matter of bad behaviour in dialogue,

including critical debate, as understood by the participants in question, all things considered.

Advises and rules of this type are left out of dialogical logic at our peril. Say good-bye to every non-pragmatic Semantics-of-Language basis for relevance theory, and shift your attention to effective discussion!

An advice of the form “Say only what is relevant” is obviously less useful for our purposes. A maxim “Say nothing which you believe to be untrue”, if incorporated among our basic norms for DL, may discourage uses of moves corresponding to the argument form *reductio ad absurdum* (see above).

16.3 Expediting the enterprise

Much of what we have discussed above can be summed up in a few lines about how to facilitate the construction and use of logic in dialogical practice. This comprises the tasks of

- facilitating *the study of counter-dialogical mentalities* by establishing this as an important discipline within the science of logic;
- facilitating *access* to thematic discussion by the same means;
- facilitating discussion as *thematic dialogue* by formulating the theme as *initial thesis* in advance;
- facilitating *acceptability of rules* by normative reconstruction of suggested sets of rules. Above we mentioned a number (not all) of “natural norms” which can be shown to be served by the rules of dialogue logic (this has not been shown here). These norms are natural in the sense that they may be expected to — after some deliberation has taken place — evoke significant global consent.
- last, but not least: facilitating the *debatability of utterances* by studying the influence on critical expediency of the choice of general non-referring particles (and perhaps introducing formal$_2$ rules for their uses).

16.4 Cooperation

To say, as some do, that to introduce logic into human dialogue means to construct “weapons” in which, tragically, other people become losers is, as we have tried to show, a dangerous attitude. Similarly the notion of a discussion with exactly two parties makes some people hesitant. They do not seem to appreciate the fact that proponent and opponent *together* probe into a statement’s tenability under critical exploration — whether the statement has been uttered tentatively or dogmatically.

Cooperation takes many forms, each form more or less well adapted to one of the many ends for which humans may be prepared to cooperate. To identify a conflict of opinions and subsequently arranging a discussion according to agreed

rules constitutes a mature form of democratic cooperation. This is not to say that conflicts of opinion and dialogical logic could not be developed as *discussions among more than two parties.*[40] However, definitions and rules would have to be carefully worked out — a task for the future.

16.5 *Real conventions, or the enhancement of logical understanding through international standardization procedures*

Finally, one ought to strive toward a small number of globally accepted packages of well-defined fundamental categories — the logical dimensions of verbal intercourse — as well as for suitable terms to express them by. Only then can one hope to end up with an effective dialogical logic with a wide range and firm backing.

The procedures chosen should be such that they issue in social contracts, i.e., in *real conventions on logical fundamentals*. The problem may be approached in the same way as the no less formidable problem in the physical and technical sciences, whose scientists have for years cooperated in selecting and defining fundamental dimensions and recommending their use in practice. They have done this, for all branches of these sciences, by means of national and thereafter international standardization committees.

"Real logical conventions" alludes to this procedure, which practice-minded logicians ought to scrutinize. Committee members should be highly qualified scholars and specialists from all possible branches of logic. Agreement can be reached through exposition, first, of the problems which a sought for category is intended to solve — e.g., how to describe and argue about change; second, by an exposition of the historical suggestions in this respect, their relative strengths and weaknesses. In other words, of the *problem-solving value* of the categories recommended as basic. This problem-solving value has objective features; in all other respects ontology is left behind.

Once the choice of fundamental and well-defined categories for logical verbal intercourse is made, the committee may go on to other problematic aspects of a dialogically serviceable logic. Now also basic formal$_2$ rules defining the use of these categories in two-role uses of language can be formulated.

The categorial standardizations are now — temporarily, at least — fundamental and the rules *conventionally valid* in and for the broad *company* (Crawshay-Williams) of language users which the members of the committee can be said to represent. The standards are allowed this logical status as a consequence of the competence of the international committee members, which is controllable, combined with the knowledge that these persons have seriously debated the question between them, paying attention to external comments in the media.

Such standardizations are *revocably fundamental*, and *revocably valid*. At a later time they may be reconsidered and replaced through new recommendations

[40]Nor does it mean that the field could not somehow be enriched with additional rules for enhancing the growth of "coalescence" [Gilbert, 1994], but the need for them do not seem nearly as pressing as Gilbert seems to think.

resulting from the same procedures. This has for a long time been the situation in physics, where an "anti-autocratic decisionism", as we might call it, has worked very well. (The term is coined and introduced here in contrast to the autocratic "decisionism" of Carl Schmitt, the Nazi philosopher of law in Hitler's service.)

The conventions that logicians and other philosophers have referred to so far are semi-conventions — non-explicit conventions hidden in some Western language or uses of language. They are only of *semi-conventional validity.*

Notice that one can very well have several systems of fundamental logical dimensions and terms next to one another, each one rooted in a different philosophy on what should be basic in logic, and why. When this becomes the situation in logic, this means that texts based on one system of logical dimensions will be sufficiently understandable to supporters of another system to enable them to discuss arguments based on the first system with supporters of that system, and vice-versa. This implies that supporters of different systems can also discuss amendments with one another. Little by little the need for communicative and discussion-supporting logics will be recognized in wider circles, and the differences between logics in actual practice will flatten out.

BIBLIOGRAPHY

See also the bibliography to Barth 1992.

[Albert, 1991] H. Albert. *Treatise on Critical Reason*. Princeton University Press, Princeton, NJ, 1985, 1991.

[Albert, 1996] H. Albert. *Der Mythos des Rahmens und der moderne Antirealismus. Zur Kritik des idealistischen R̀uckfalls im gegenwärtigen Denkens*. Forum for European Culture (First Sir Karl Popper lecture), Amsterdam, 1996.

[Apostel, 1972] L. Apostel. Negation: the tension between ontological positivity (negationless positivity) and Anthropological negativity (positively described), *Logique et Analyse*, 15, 209–317, 1972.

[Apostel *et al.*, 1972] L. Apostel, J. Buelens, M. de Mey, L De Ryck-Tasmowski, F. Vandamme and F. Van Dun, eds. *Logique et Analyse* 15, 1–401 (Nos. 57–59, special issue on negation), 1972.

[Barth, 1972] E. M. Barth. *Evaluaties*, Van Gorcum & Comp, Assen, 1972.

[Barth, 1975] E. M. Barth. Problem-solving validity and semi-conventional validity in formal logic. In *Proceedings of 5th ICLMPS*, London, Ontario, 1975.

[Barth, 1977] E. M. Barth. Commentaar op en advies voor een leerplan logica, *Algemeen Nederlands Tijdschrift voor Wijsbegeerte*, 69, 66–180, 1977.

[Barth, 1979] E. M. Barth. *Perspectives on Analytic Philosophy*. Meded. Kon. Ned. Akad. van Wet. N.R. 42, No. 2. Amsterdam: North-Holland, 1979.

[Barth, 1981] E. M. Barth. *The Logic of the Articles in Traditional Philosophy. A Contribution to the Study of Conceptual Structures* 2nd edition. Kluwer Academic Publishers, Dordrecht', 1981.

[Barth, 1982] E. M. Barth. A normative-pragmatical foundation of the rules of some systems of formal$_3$ dialectics. In [Barth and Martens, 1982, pp. 159–170].

[Barth, 1983] E. M. Barth. A two-role problem-analytic analysis of Montague's PTQ. (paper 7th ICLMPS, Salzburg), 1983.

[Barth, 1985a] E. M. Barth. Toward a praxis-oriented Theory of Argumentation. In *Dialogue - An Interdisciplinary Approach*, M. Dascal, ed. pp. 73-86. Amsterdam/Philadelphia: John Benjamins, 1985. Transl. as: "Para uma teoria da argumentaa̧o baseada na práxis", in: Cadernos de Estudos Linguísticos, Universitade Estadual de Campinas, Departamento de Lingüística, pp. 69–80 1986.

[Barth, 1985b] E. M. Barth. Two-phase epistemology and models for dialogue logic. In *Recent developments in dialogue logics*, J.P. van Bendegem, ed. Philosophica 35, 69–88, 1985.

[Barth, 1985c] E. M. Barth. A new field: Empirical logic. Bioprograms, logemes, and logics as institutions. In Walton, ed. pp. 375–388, 1985. Repr. in [Barth *et al.*, 1992].

[Barth, 1985d] E. M. Barth. Dialectical fields and transformations: Brouwer fields, Beth fields, and Naess transformations. *Philosophia naturalis*, 21, P. Weingartner and C. Pühringer, eds. pp. 425–434, 1985.

[Barth, 1987] E. M. Barth. Contradictions and symmetry rage in the logical Interregnum. An essay in Empirical Logic. In *Logos and Pragma. Essays on the Philosophy of Language in Honour of Professor Gabriel Nuchelmans*, L.M. de Rijk and H.A.G. Braakhuis, eds. pp. 215–239. Ingenium Publ., Nijmegen, 1987.

[Barth, 1990] E. M. Barth. In the service of human society: Formal, informal or anti-logical? The philosophy of the logician E.W. Beth, *Informal Logic* XII, 1–10, 1990.

[Barth, 1992] E. M. Barth. Logic as a tool in the analysis of political cognition and argument. In [Barth and Krabbe, 1992, pp. 151–169].

[Barth, 1992a] E. M. Barth. Dialogical approaches. In: [Dascal *et al.*, 1992, 1996], pp. 663–676, 1992. Bibliography: cf. (idem) 1995.

[Barth, 1993] E. M. Barth. Systems of Logical Representation and Inference: An Empiricist Approach to Cognitive Science. In Jorna a.o., eds. pp. 48–65, 1993.

[Barth, 1995] E. M. Barth. A categorial positioning of contemporary uses of generic terms is impossible. In *Festschrift for Professor Marian Przelecki*, J.J. Jadacki and M. Omyla, eds. 1995.

[Barth, 1996] E. M. Barth. Logeme Shifts and the Growth of Pragmatism. In *Pragmatisme et phénoménologie dans leurs relations avec les sciences formelles* (Actes du Congrès Internat. Henri Poincaré, Nancy 1994.) Philosophia Scientiae. Travaux d'histoire et de philosophie des sciences 1, 91–108, 1996.

[Barth and Krabbe, 1978] E. M. Barth and E.C.W. Krabbe. Formele$_3$ dialektiek: instrumenten ter beslechting van conflicten over geuite meningen, *Spektator*, 7, F.H. v. Eemeren a.o., eds. pp. 307–341, 1978.

[Barth and Krabbe, 1982] E. M. Barth and E. C. W. Krabbe. *From Axiom to Dialogue - A Philosophical Study of Logics and Argumentation*. Walter de Gruyter, Berlin/New York, 1982.

[Barth and Krabbe, 1992] E. M. Barth and E. C. W. Krabbe, eds. *Logic and Political Culture* , (Proc. Coll. Logic and Politics, Amsterdam 1990. Kon. Ned. Akad. van Wet., N.R., deel 149. Amsterdam etc.: North-Holland, 1982.

[Barth and Martens, 1977] E. M. Barth and J.L. Martens. Argumentum ad hominem: From chaos to formal dialectic. The dialogical tableau method as a tool in the theory of fallacy, *Analyse* 20, 76–96, 1977.

[Barth and Martens, 1982] E. M. Barth and J.L. Martens, eds. *Argumentation: Approaches to Theory Formation*. Proc. Groningen Conf. on the Theory of Argumentation, Oct.1978. Chron. bibliogr. A. van Hoof. Amsterdam: John Benjamins, 1982.

[Barth and Wiche, 1986] E. M. Barth and R. Wiche, eds. *Problems, Functions, and Semantic Roles. A Pragmatists' Analysis of Montague's Theory of Sentence Meaning*. Walter de Gruyter, Berlin/New York, 1986.

[Barth *et al.*, 1992] E. M. Barth, J. Vandormael and F. Vandamme, eds. *From an Empirical Point of View - The Empirical Turn in Logic*, Ghent, 1992.

[Barth *et al.*, 1999] E. M. Barth, G. Heinzman, H. Visser, eds. *Un Logicien consciencieux. La philosophie de Evert Willem Beth / A Conscientious Logician. The Philosophy of Evert Willem Beth*. In Philosophiae Scientiae 3, No. 4. Paris: Éditions Kimé, 1999.

[Bendegem, 1982] J. P. van Bendegem. Pragmatics and mathematics, or, how do mathematicians talk? *Philosophica*, 29, 97–118, 1982.

[Bendegem, 1983] J. P. van Bendegem. The Return to Empirical Mathematics, in *Representation Use and Acquisition of Knowledge or the Manufacture of Knowledge*, F. Vandamme, ed. pp. 101–108, Ass. Int. de Cybernetique, Namur 1983.

[Bendegem, 1985] J. P. van Bendegem. *Finite, Empirical Mathematics: Outline of a Model*. Preprints from Rijksuniversiteit Gent, 1985.

[Bendegem, 1985a] J. P. van Bendegem. Dialogue logic and problem solving, in [Bendegem, 1985b], pp. 113–134, 1985.

[Bendegem, 1985b] J. P. van Bendegem, ed. *Recent Developments in Dialogue Logic*, special issue of *Philosophica* (Gent), 35, 1985.

[Bendegem, 1988] J. P. van Bendegem. Non-formal properties of real mathematical proofs. In Arthur Fine and Jarrett Leplin, eds., *PSA 1988*. Volume One, PSA, East Lansing, pp. 249-254, 1988.

[Bendegem, 1993] J. P. van Bendegem. Real-life mathematics versus ideal mathematics: The ugly truth, in [Krabbe *et al.*, 1993], pp. 263–272, 1993.

[Benthem van den Bergh, 1980] G. van Benthem van den Bergh. *De Staat van Geweld: Essays*, Amsterdam: Meulenhoff, 1980.

[Beth, 1959a] E. W. Beth. *The Foundations of Mathematics. A Study in the Philosophy of Science*. Amsterdam: North-Holland, 1959; 2nd ed. 1964. 3rd ed. New York: Harper & Row 1966.

[Beth, 1959b] E. W. Beth. Het deductie-theoretisch analogon van de methode der semantische tableaux, lecture given 1958 before the Dutch Assoc. for Logic and the Philosophy of Science; see *Inventory*, p. 87, 1995.

[Beth, 1962] E. W. Beth. *Formal Methods. An Introduction to Symbolic Logic and to the Study of Effective Operations in Arithmetic and Logic*, Doredrecht: Reidel, 1962.

[Beth, 1968] E. W. Beth. *Science a Road to Wisdom: Collected Philosophical Studies* Dordrecht: Reidel (Kluwer), 1968. (Tr. by P.G.E. Wesly of Door wetenschap tot wijsheid, 1964).

[Beth, 1970] E. W. Beth. *Aspects of Modern Logic*, D. Reidel, Dordrecht, 1970. Trans. of E. W. Beth, *Moderne logica*, 1967.

[Beth, 1995] E. W. Beth. *Inventory*, 1995. See Velthuys-Bechthold, P.J.M.

[Boyer, 1959] C. B. Boyer. *The History of the Calculus and Its Conceptual Development*, Dover Publications, New York, 1959 (1939).

[Buber, 1923] M. Buber. *Ich und Du*, Inzel-Verlag, Leipzig, 1923.

[Cassirer, 1910] E. Cassirer. *Substanzbegriff und Funktionsbegriff*, Berlin. Tr. as Substance and Function (1923), 1910.

[Cornford, 1957] F. M. Cornford. *From Religion to Philosophy. A Study in the Origins of Western Speculation*. Harper and Row, New York, 1957 (1912).

[Crawshay-Williams, 1957] R. Crawshay-Williams. *Methods and Criteria of Reasoning. An Inquiry into the Structure of Controversy*, London: Routledge and Kegan Paul, 1957.

[Dascal *et al.*, 1992, 1996] M. Dascal, D. Gerhardus, K. Lorenz and G. Meggle, eds. *Sprachphilosophie – Philosophy of Language – La philosophie du language. An International Handbook of Contemporary Research I-II*. Walter de Gruyter, Berlin/New York, 1992. Bibliography, idem 1995.

[Dascal *et al.*, 1996] M. Dascal, J. Hintikka and K.Lorenz. Jeux dans le langage/Games in langauge/Spiel ider Sprach. In [Dascal *et al.*, 1992, 1996], 1996.

[Davis and Hersh, 1972] M. Davis and R. Hersh Non-Standard Analysis, *Scientific American*, 226:78–86, 1972. Repr., P.J. Davis and R. Hersh, *The Mathematical Experience*, 1981.

[Drieschner, 1963] R. Drieschner. *Untersuchung zur dialogischen Deutung der Logik* (diss.), 1963.

[Ellerman, 1988] D. P. Ellerman. Category Theory and Concrete Universals, *Erkenntnis* 28:409–429, 1988.

[Fine, 1985] K. Fine. *Reasoning With Arbitrary Objects*. Oxford: Blackwell, Arist. Soc. Ser., Vol. 3, 1985.

[Finocchiaro, 1980] M. A. Finocchiaro. *Galileo and the Art of Reasoning - Rhetorical Foundations of Logic and Scientific Method*. Dordrecht: Reidel (Kluwer), 1980.

[Finocchiaro, 1992] M. A. Finocchiaro. Logic, Politics, and Gramsci, in [Barth and Krabbe, 1978], 1992.

[Freytag-Löringhoff, 1955] B. Baron von Freytag-Löringhoff. *Logik. Ihr System und ihr Verhältnis zur Logistik*, Stuttgart: Verlag W. Kohlhammer, 1955.

[Gaukroger, 1995] S. Gaukroger. *Descartes. An Intellectual Biography*, Oxford: Clarendon Press, 1995.

[Geach, 1968] P. T. Geach. *Reference and Generality. An Examination of Some Medieval and Modern Theories*, revised ed. Ithaca, NY, 1968.

[Geach, 1972] P. T. Geach. *Logic Matters*, Oxford, 1972.

[Gethmann, 1982] C. F. Gethman, ed. *Logik und Pragmatik.**Z**um Rechtfertigungsproblem logischer Sprachregeln*, Frankfurt am Main: Suhrkamp, 1982.

[Gilbert, 1994] M. Gilbert. Feminism, Argumentation and Coalescence, *Formal Logic*XVI, 95-113, 1994.

[Glasersfeld, 1998] E. von Glasersfeld. Die Radikal-Konstruktivistische Wissenstheorie, *Ethik und Sozialwissenschaften [EuS]*, 9, 503–511, 1998.

[Gullvåg and Naess, 1996] I. Gullvåg and A. Naess. Vagueness and Ambiguity. In [Dascal *et al.*, 1992, 1996].

[Haack, 1974] S. Haack. *Deviant Logic*. Cambridge: Cambridge UP, 1974.

[Hacking, 1975] I. Hacking. *Why Does Language Matter to Philosophy?* Cambridge: Cambridge UP, 1975.

[Hamblin, 1970] C. L. Hamblin. *Fallacies*. London: Methuen, 1970.

[Harman, 1984] G. Harman. Logic and reasoning, *Synthese* 60, 107–127, 1984.

[Heyting, 1956] A. Heyting. *Intuitionism, an Introduction*. Amsterdam: North-Holland, 1956.

[Hintikka, 1968] J. Hintikka. Language games for quantifiers, in: *Studies in Logical Theory*, Am. Philos. Quart. Monogr. Ser., No. 2, 1968.

[Hintikka, 1979] J. Hintikka. Quantifiers vs. quantification theory. In *Game-theoretical Semantics*, E. Saarinen, ed. pp. 49–79. D. Reidel, Dordrecht, 1979.

[Hintikka, 1979] J. Hintikka. Quantifiers in natural logic: some logical problems. In *Game-theoretical Semantics*, E. Saarinen, ed. pp. 81–117, Reidel, Dordrecht, 1979.

[Hoepelman and van Hoof, 1988] J. Ph. Hoepelman and A. J. M. van Hoof. The success of failure. The concept of failure in Dialogue Logics and its relevance for natural language semantics, in: D. Vargha (ed.), *Coling Budapest. Proc. 12th Int. Conf. on Comput. Linguistics*, Budapest; 250–255, 1988.

[Hoepelman and van Hoof, 1993] J. Ph. Hoepelman and A.J.M. van Hoof. Default and dogma, in: [Krabbe *et al.*, 1993], pp. 141-158, 1993.

[Hoof, 1995] A. J. M. van Hoof. *Der logisch-pragmatischen Gebrauch von Konditionalsätzen. Eine dialog-logische Analyse*. Berlin, Springer, 1995.

[Hoof and Hoepelman, 1991] A. J. M. van Hoof and J. Hoepelman Non-monotonicity and conditionals in Dialogue Logic. In: Schroeder-Heiser, P. (ed.), *Extensions of Logic Programming. Proc. Int. Workshop Tübingen*, FRG, Dec. 1989. Berlin etc.: Springer, 1991.

[Inhetveen, 1982] R. Inhetveen. Ein konstruktiver Weg zur Semantik der 'Möglichen Welten'. In: [Barth and Martens, 1982], pp. 133–141, 1982.

[Janich, 1992] P. Janich, ed. *Entwicklungen der methodischen Philosophie*, Frankfurt a/M: Suhrkampf, 1992.

[Janich, 1998] P. Janich. Radikal halbherzig. Die Wissenstheorie E. v. Glaserfelds, *Ethik und Sozialwissenschaften – Streitforum für Erwägungskultur [EuS]* 9, 535–537, 1998.

[Jaspers, 1967] K. Jaspers. *Drei Gründer des Philosopherens: Plato, Augustin, Kant*, R. Piper, Munich, 1967 (1957).

[Johnson, 1993] R. Johnson. Dialectical fields and manifest rationality, in: [Krabbe *et al.*, 1993, pp. 201–210].

[Johnson and Blair, 1994] R. Johnson and J.A. Blair. *Logical Self-Defence*. 1st ed. 1977. New York etc.: McGraw-Hill Ryerson, 1994.

[Johnstone, 1959] H. W. Johnstone Jr. *Philosphy and Argument*, Pennsylvania State University Press, 1959.

[Jorna, 1993] R. Jorna, B. van Heusden and R. Posner (eds.) *Signs, Search and Communication*. Berlin: Walter de Gruyter, 1993.

[Kiesewetter, 1996] H. Kiesewetter. *Von Hegel zu Hitler. Die politische Verwirklichung einer totalitären Machtsstaatstheorie in Deutschland (1815-1945)* (2nd ed., revised, extended). Frankfurt am Main: Peter Lang, 1996.

[Kitcher, 1982] P. Kitcher. *The Nature of Mathematical Knowledge*. Oxford etc.: Oxford UP, 1982.

[Kitcher, 1986] P. Kitcher. *Vaulting Ambition*, Cambridge, Mass: MIT Press, 1986.

[Krabbe, 1978] E. C. W. Krabbe. The adequacy of material dialogue-games, *Notre Dame J. of Formal Logic* 19, 321–330, 1978.

[Krabbe, 1982] E. C. W. Krabbe. *Studies in Dialogical Logic* (diss.), Groningen: Rijksuniversiteit Groningen, 1982.

[Krabbe, 1985] E. C. W. Krabbe. Non-cumulative dialectical models and formal dialectics, *J. Philos. Logic* 14, 129–168, 1985.

[Krabbe, 1986] E. C. W. Krabbe. A theory of modal dialectics, *J. Philos. Logic* 15, 191–217, 1986.

[Krabbe *et al.*, 1993] E. C. W. Krabbe, R. J. Dalitz and P. A. Smit, eds. *Empirical Logic and Public Debate*, Amsterdam/Atlanta, GA: Rodopi, 1993.

[Lenk, 1968] H. Lenk. *Kritik der logischen Konstanten. Philosophische Begründungen der Urteilsformen vom Idealismus bis zur Gegenwart* (A Critique of the Logical Constants. Philos. Foundations of the Forms of Judgement from Idealism to the Present), Berlin. Rev., enl. ed. of post-doctoral thesis, 1966.

[Lorenz, 1968] K. Lorenz. Dialogspiele als semantische Grundlage von Logikalkülen I- II, *Archif für mathem. Logik u. Grundl.forschung* 11, 32–55, 73–100, 1968.

[Lorenz, 1971] K. Lorenz. *Elemente der Sprachkritik - Eine Alternative zum Dogmatismus und Skeptizismus in der Analytischen Philosophie, vol. II*. Frankfurt/M.: Suhrkamp. 1st ed. 1970, 1971.

[Lorenz, 1973] K. Lorenz. Rules versus theorems, *J. Philos. Logic* 2, 352–369, 1973.

[Lorenz, 1980] K. Lorenz. Logik, dialogische, in: *Hist. Wörterbuch der Philos*. (eds. J. Ritter, K. Gründer). Basel etc.: Schwabe. Vol V, 402–411, 1980.

[Lorenz, 1980a] K. Lorenz. Main ideas of Dialogic Logic, in: *Studies in Logic, Grammar, and Rhetoric I*, 17–34, 1980.

[Lorenz, 1980b] K. Lorenz. Dialogic logic, in: W. Marciszewski (ed.), *Dictionary of Logics as Applied in the study of Language. Concepts – Methods – Theories*, The Hague; 117–125, 1980.

[Lorenz, 1982] K. Lorenz. On the criteria for the choice of the rules of dialogic logic, in[Barth and Martens, 1982], pp. 145–157, 1982.

[Lorenz, 1982a] K. Lorenz. Die irreführende Gleichsetzung von Begründungen und Argumentationen. Bemerkungen zu einem monologischen Missverständnis in der dialogischen Logik, in:[Gethmann, 1982], pp. 78–91, 1982.

[Lorenz, 1992] K. Lorenz. Das dialogische Prinzip in der Philosophie. Ein Beitrag zurÜberwindung der Subjekt-Objekt-Spaltung, in: [Janich, 1992], 1992.

[Lorenzen, 1960] P. Lorenzen. Logik und Agon, congress presentation; in: Atti del XII Congresso Internationale di Filosofia (Venice 1958), Firenze: Sansoni; repr. in[Lorenzen and Lorenz, 1978], 1960.

[Lorenzen, 1969] P. Lorenzen. *Normative Logic and Ethics*, Mannheim/Zürich: Bibliographisches Institut, 1969.

[Lorenzen, 1982] P. Lorenzen. Die dialogische Begründung von Logikkalkülen. In: [Barth and Martens, 1982], pp. 23–54, 1982.

[Lorenzen, 1987] P. Lorenzen. *Critique of Political and Technical Reason*. The Evert Willem Beth Lectures 1980, publ. as *Synthese* 71 No. 2 (127–218), 1987.

[Lorenzen and Lorenz, 1978] P. Lorenzen and K. Lorenz. *Dialogische Logik*. Darmstadt: Wissensch. Buchgesellschaft, 1978.

[Mackenzie, 1979] J. D. Mackenzie. Question-Begging in Non-Cumulative Systems, *Journal of Philosophical Logic* 8, 117–133, 1979.

[Mackenzie, 1989] J. D. Mackenzie. Reasoning and logic, *Synthese* 79, 99–117, 1989.

[Mackenzie, 1990] J. D. Mackenzie. Four dialogue systems, *Studia logica* XLIX, 567–583, 1990.

[Mackenzie and Priest, 1990] J. D. Mackenzie and G. Priest. Paraconsistent dialogues, or, how to start talking to Cretans, *Logique et Analyse* 131–132, 339–357, 1990.

[Manin, 1977] Yu. I. Manin. *A course in Mathematical Logic*. New York/Heidelberg/Berlin: Springer-Verlag. Orig. Russian ed. 1974, 1977.

[Maritain, 1933] J. Maritain. *Éléments de philosophie II : L'ordre des concepts. 1: Petite logique (logique formelle)*. Paris: Tegui, 1923. (11th impr.) 1933.

[Mittelstaedt, 1978] P. Mittelstaedt. *Quantum Logic*. Dordrecht: D. Reidel, 1978.

[Mittelstaedt, 1997] P. Mittelstaedt. Am Anfang war eben nicht der Garten, in *Ethik und Sozialwissenschaften – Streitforum für Erwägungskultur* [EuS] 8, 1997.

[Moore and Hobbes, 1996] D. Moore and D. Hobbes. Computational Uses of Philosophical Dialogue Theories, *Informal Logic* 18, 2/3, 1996.

[Morris, 1938] C. Morris. *Foundations of the Theory of Signs*. Chicago, 1938.

[Naess, 1936] A. Naess. *Erkenntnis und wissenschaftliches Verhalten*. Eng. tr. forthcoming in Arne Naess: Collected Works (ed. Glaser), 1936.

[Naess, 1947] A. Naess. *End del elementære logiske emner*, Universitetsforlaget, Oslo, 1959 (and later editions). Translated as [Naess, 1981]. German translation 1975, Dutch translation 1977.

[Naess, 1953] A. Naess. *Interpretation and Preciseness, A Contribution to the Theory of Communication*, Oslo, 1953. 2nd ed. 1996.

[Naess, 1956] A. Naess. *Wie Fördert man heute die empirische Bewegung? Eine Auseinandersetzung mit dem Empirismus von Otto Neurath und Rudolph Carnap*, Institute for Philosohy and the History of Ideas, Oslo University, 1956 (1939).

[Naess, 1981] A. Naess. Communication and Argument – Elements of Applied Semantics (or. ed. c. 1940). Oslo etc.: Universitetsforlaget, 1981 (1966).

[Naess, 1992a] A. Naess. Arguing under deep disagreement. Contains crucial selections from 1992b. In: [Barth and Krabbe, 1992], pp. 123–131, 1992.

[Naess, 1992b] A. Naess. How can the empirical movement be promoted today? A discussion of the empiricism of Otto Neurath and Rudolph Carnap (complete tr. of [Naess, 1956]). In: [Barth *et al.*, 1992], pp. 107–155, 1992.

[Naess, 1993] A. Naess. You assert this? An empirical study of weight expressions, in: [Krabbe *et al.*, 1993], 121–132, 1993.

[Ong, 1983] W. J. Ong. *Ramus, Method, and the Decay of dialogue. From the Art of Discourse to the Art of Reason*. Harvard University Press, Cambridge, MA, 1983 (1958).

[Paalman de Miranda, 1990] A. B. Paalman de Miranda. Note on a paper of D.P. Ellerman, *Erkenntnis* 33, 397f, 1990.

[Popper, 1959] K. Popper. *The Logic of Scientific Discovery* (tr. from *Logik der Forschung*, 1934). London: Hutchinson, 1959.

[Prior, 1977] A. N. Prior. Egocentric logic. In: A. N. Prior and K. Fine, *Worlds, Times and Selves*. London: Duckworth; pp. 28–50, 1977.

[Rahman and Carnielli, 1998] S. Rahman and W. Carnielli. The Dialogical Approach to Paraconsistency, Bericht Nr. 8, Dept. of Philos., Univ. des Saarlandes, Saarbrücken. 33 pp, 1998.

[Rescher, 1977] N. Rescher. *Dialectics: A Controversy-Oriented Approach to the Theory of Knowledge*. Albany, N.Y.: State Univ. of N.Y. Press, 1977.

[Rose, 1990] P. L. Rose. *Revolutionary Antisemitism in Germany - From Kant to Wagner*, Princeton, N.J.: Princeton U.P, 1990.

[Scholz, 1967] H. Scholz. *Abriss der Geschichte der Logic*. Karl Alber, Freiburg and Munich, 1967 (1931).

[Smit, 1996] P. A. Smit. *Lenins Politieke Logica. Een analyse van zijn polemieken, voorstellingswereld, politieke tactieken en strategieen*, Tilburg: Tilburg University Press, 1996.

[Smit, 1993] P. A. Smit. The logic of Virtue and Terror, in: [Krabbe *et al.*, 1993], pp. 239–251, 1993.

[Stigt, 1990] W. P. van Stigt. *Brouwer's Intuitionism*, Amsterdam etc.: North-Holland, 1990.

[Tamminga, 1994] A. Tamminga. Logics and rejection, *Logique et analyse* 146, 169–207, 1994.

[Taschner, 1998] R. Taschner. Der Blick in Gottes Karten, *Ethik und Sozialwissenschaften – Streitforum für Erwägungskultur [EuS]* 9, 571–573, 1998.

[Theunissen, 1977] M. Theunissen. *Der Andere - Studien zur Sozialontologie der Gegenwart*. Berlin: Walter de Gruyter, 1977 (1965).

[Toulmin, 1958] S. Toulmin. *The Uses of Argument*. Cambridge: Cambridge U.P, 1958.2nd ed. 1964.

[Velthuys-Bechthold, 1995] P. J. M. Velthuys-Bechthold. *Inventory of Evert Willem Beth (1908–1964)*, Haarlem: Rijksarchief Noord-Holland, 1995.

[Walton, 1985a] D. Walton, ed. The Logic of Dialogue, *Synthese* 63 No. 3, 1985.

[Walton, 1985b] D. Walton. *Arguer's Position. A Pragmatic Study of Ad Hominem Attack, Criticism, Refutation, and Fallacy*, Westport, Connecticut and London, England: Greenwood Press, 1985.

[Walton and Krabbe, 1995] D. Walton and E. C. W. Krabbe. *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*, Albany, NY: State University of New York Press, 1995.

[Weber, 1998] J. Weber. Angepasste Monologe? *Ethik und Sozialwissenschaften – Streitforum für Erwägungskultur [EuS]* 9, 577–579, 1998.

[Wiche, 1993] R. Wiche. Gerrit Mannoury on the communicative functions of negations in ordinary language, in [Krabbe *et al.*, 1993], pp. 133–140, 1993.

[Wiche, 1999] R. Wiche. Dialogische Theorie und dialogische Praxis, *Ethik und Sozialwissenschaften – Streitforum für Erwägungskultur [EuS]* 10, 654f, 1999.

[Woods, 1974] J. Woods. *The Logic of Fiction*, The Hague and Paris: Mouton, 1974.

[Woods, 1992] J. Woods. Public Policy and Standoffs of Force Five, in [Barth and Krabbe, 1992], pp. 97–108, 1992. Reprinted in [Woods, 2001].

[Woods, 1993] J. Woods. Dialectical blindspots. *Philosophy and Rhetoric*, 26, 251–265, 1993.

[Woods, 1994] J. Woods. Is the Theoretical Unity of the Fallacies Possible? *Informal Logic* XVI, 77–85, 1994. Reprinted in [Woods, 2001].

[Woods, 2001] J. Woods. *The Death of Argument: Fallacies and Other Seductions*, Vale Press, Newport News, 2001, to appear.

[Woods and Walton, 1978] J. Woods and D. Walton. Arresting circles in formal dialectic. *Journal of Philosophical Loigc*, 7, 73–90, 1978.

[Woods and Walton, 1982] J. Woods and D. Walton. *Argument: The Logic of the Fallacies*, Toronto and New York: McGraw-Hill Ryerson, 1982.

[Woods and Walton, 1989] J. Woods and D. Walton. *Fallacies: Selected Papers 1972–1982*. Foris/de Gruyter, Berlin and New York, 1989.

[Zahn, 1998] P. Zahn. Explikations-Anforderungen an den Radikalen Konstruktivismus, *Ethik und Sozialwissenschaften – Streitforum für Erwägungskultur [EuS]* 9, 579–581, 1998.

JAAKKO HINTIKKA, ILPO HALONEN AND ARTO MUTANEN

INTERROGATIVE LOGIC AS A GENERAL THEORY OF REASONING

1 THE GENERALITY OF INTERROGATIVE REASONING

The interrogative approach to reasoning and argumentation is not just one approach to its subject among many. Both historically and systematically, it is arguably the first and foremost theory of reasoning.

Historically, the first model of reasoning to capture philosophers' attention was the Socratic method of questioning or *elenchus*. Plato systematized it into a practice of questioning games used in the Academy to train his students in philosophical and scientific reasoning. Aristotle developed in the *Topica* a theory of such interrogative argumentation, emphasizing among other things its usefulness for scientific purposes. In so doing, his attention was caught by answers that are necessitated by earlier ones or, as we would express it, are logically implied by them. His study of such answers became the first study of logic in history. (Cf. here [Hintikka, 1996a].)

Thus the interrogative method is historically speaking not just one way of approaching logic and argumentation. It is *the* theory of reasoning, logical as well as empirical, comprising deductive logic as a special case.

Systematically, too, the interrogative approach can be argued to be a general theory of reasoning. In order to see the sense in which this is the case, consider someone's line of reasoning, expressed in the form of a series of propositions (sentences). What is a new proposition in such a series like? Either the information it codifies is contained in the earlier propositions in the series, or it is not. In the former case, the step is a logical (deductive) consequence of these earlier steps. Such argumentative steps are to be studied by means of deductive logic.

In the latter case, if the arguer is rational, he or she must know where the new information comes from. Otherwise we are not dealing with reasoning and inference but mere guesswork. We will use as a technical term for all sources of information the word *oracle*. Since the information is new, the inquirer must somehow have received this information as a response to his or her own initiative, which is an action directed to some particular oracle. Since this source of information is therefore known, we might as well think of the new information as an answer to a question the inquirer has addressed to the oracle in question. But what this means is that such rational line of reasoning can in principle be thought of as an interrogative process. The Greeks were right: all reasoning can be thought of as questioning. The main

superficial change as compared with Socratic *elenchus* is that purely logical (deductive) steps are in our treatment distinguished from interrogative ones. This modification was anticipated by Aristotle, even though it is not realized clearly enough by most historians that he still considered even syllogistic inferences in science merely as special kinds of question-answer steps in interrogative inquiry.

2 THE BASIC RULES OF INTERROGATIVE LOGIC

What, then, is the logic (structure) of interrogative inquiry? We will first consider a simple case which nevertheless brings out several important features of reasoning and inquiry in general. This case is characterized by the following features:

1. There is only one oracle.

2. The set of answers the oracle will provide remains constant throughout the inquiry.

3. All of the oracle's answers are true, and known by the inquirer to be true.

Given these assumptions, the logic of interrogative reasoning will at first sight seem exceedingly simple. We can formulate the rules for such a procedure in the form of rules for a game played by the inquirer against an oracle (or a number of oracles). The book-keeping method we use is a variant of Beth's [1955] *tableau* method. All the moves are initiated by the inquirer. The only role that the oracle plays is to respond to the inquirer's questions. There are certain initial premises on the left side of the *tableau* and the proposition to be interrogatively established (proved) on the right side. There are two kinds of moves, logical inference moves and interrogative moves. The logical inference moves are simply a variant of the *tableau*-building rules of the usual *tableau* method. For simplicity, we will here and in the rest of this paper assume that all the formulas we are dealing with are in the negation normal form, i.e. that all negation signs precede immediately atomic formulas or identities, unless otherwise indicated. The basic concepts of the *tableau* method, such as column, closure and *subtableau*, are used as usual. An example of the *tableau*-building rules is the following:

If $(S_1 \vee S_2)$ occurs on the left side of a *subtableau*, it is divided into two *subtableaux* which have as a new formula in their left side S_1 and S_2, respectively.

In order to obtain a more concise notation we will formulate the *tableau*-building rules as inverses of the corresponding Genzen-style (sequent) rules. The formulas on the left side of a *(sub)tableau* go to the antecedent of the

corresponding sequent, the formulas on the right side go to the consequent. Thus in this notation the disjunction rule becomes

$$\text{(L.}\vee\text{)} \qquad \frac{\Gamma, (S_1 \vee S_2) \rightarrow \Delta}{\Gamma, S_1 \rightarrow \Delta \quad \Gamma, S_2 \rightarrow \Delta}\,.$$

The other rules can be formulated as follows:

$$\text{(L.\&)} \qquad \frac{\Gamma, (S_1 \& S_2) \rightarrow \Delta}{\Gamma, S_1, S_2 \rightarrow \Delta}\,,$$

$$\text{(L.e)} \qquad \frac{\Gamma, (\exists x) S[x] \rightarrow \Delta}{\Gamma, S[b] \rightarrow \Delta}$$

where b is a new individual constant. Constants introduced by this rule are called *dummy names.*

$$\text{(L.a)} \qquad \frac{\Gamma, (\forall x) S[x] \rightarrow \Delta}{\Gamma, (\forall x) S[x], S[b] \rightarrow \Delta}$$

where b is an individual constant occuring in Γ, Δ or $S[x]$.

In fact, (L.a) is a special case of a more general rule:

$$\text{(L.A)} \qquad \frac{\Gamma, (\forall x) S[x] \rightarrow \Delta}{\Gamma, (\forall x) S[x], S[t] \rightarrow \Delta}$$

where t is a term built of function constants and individual constants occuring in Γ, Δ or $S[x]$.

The right-hand rules are mirror images of (L.$\vee$)–(L.A), e.g.,

$$\text{(R.}\vee\text{)} \qquad \frac{\Gamma \rightarrow \Delta, (S_1 \vee S_2)}{\Gamma \rightarrow \Delta, S_1, S_2}\,.$$

For formulas which are not in a negation normal form, we will use rules to drive negation deeper into the formulas in question, e.g.

$$\text{(L.}\sim\vee\text{)} \qquad \frac{\Gamma, \sim (S_1 \vee S_2) \rightarrow \Delta}{\Gamma, (\sim S_1 \& \sim S_2) \rightarrow \Delta}\,.$$

We also need some suitable structural rules, which allow us to permute the members on the left side or on the right side of a sequent, and closure rules. The following can serve as rules for closure:

(C.T) A tableau is closed iff all its *subtableaux* are closed.

A *subtableau* is closed if it includes one of the following:

(C.L) $\Gamma, S_1, \sim S_1 \rightarrow \Delta$

(C.R) $\Gamma \rightarrow \Delta, S_1, \sim S_1$

(C.LR) $\Gamma, S_1 \rightarrow \Delta, S_1$.

If identity is included in the underlying language, we need substitution rules. These rules are illustrated by the following examples:

(LL.=) $$\frac{\Gamma, S[a], (a = b) \rightarrow \Delta}{\Gamma, S[a], S[b], (a = b) \rightarrow \Delta},$$

(LR.=) $$\frac{\Gamma, (a = b) \rightarrow \Delta, S[a]}{\Gamma, (a = b) \rightarrow \Delta, S[a], S[b]}.$$

We also need closure rules for identity, for instance a rule that says that a *subtableau* is closed if it contains

(C.L=) $\Gamma, \sim (a = a) \rightarrow \Delta$.

For reasons that will become clear later, it is advisable to generalize (L.e) and (R.a) as to allow functional instantiation over and above the usual existential instantiation:

(L.E) $$\frac{\Gamma, S_0[(\exists x)S_1[x]] \rightarrow \Delta}{\Gamma, S_0[S_1[f(y_1, \ldots, y_n)]] \rightarrow \Delta}$$

where f is a new function constant and $(\forall y_1), \ldots, (\forall y_n)$ are all the universal quantifiers within the scope of which $(\exists x)S_1[x]$ occurs in S_0. Function constants introduced in two way are said to represent *dummy functions*.

(R.A) Analogously.

It is important to recall here and in similar rules below that S_0 is assumed to be in the negation normal form, i.e. that all negation signs in S_0 are in front of atomic formulas. (L.e) is taken to be a special case of (L.E).

Likewise it is advisable to generalize (L.$\vee$) and (R.&) an analogous way:

(L. $\vee f$) $$\frac{\Gamma, S_0[(S_1 \vee S_2)] \rightarrow \Delta}{\Gamma, S_0[(S_1 \& f(y_1, \ldots, y_n) = 0) \vee (S_2 \& f(y_1, \ldots, y_n) \neq 0)] \rightarrow \Delta}$$

Here f is a new function constant (dummy function) and $(\forall y_1), \ldots, (\forall y_n)$ are all the universal quantifiers within the scope of which $(S_1 \vee S_2)$ occurs in S_0.

(R.& f) Analogously.

The old rules (L.$\vee$) and (R.&) are taken to be special cases of the new rules (L.$\vee f$) and (R.& f).

3 RULES FOR QUESTIONING

The rule for questioning can be formulated as follows:

(L.Q) If the presupposition of a question occurs on the left side of a *subtableau*, the inquirer may address the corresponding question to the oracle. If the oracle answers, the answer is added to the left side of the *subtableau*.

For obvious reasons, there is no right-side questioning rule. Equally obviously, no dummy names must occur in an answer.

The rule (L.Q) refers to the notion of presupposition. Hence our next task is to explain this notion. We will at the same time define (tentatively) another basic notion from the theory and practice of questions, the notion of answer. In so doing, we have to distinguish different kinds of questions from each other. In a propositional question, the *presupposition* is a disjunction

$$(S_1 \vee \ldots \vee S_n)$$

and the possible *answers* are $S_1, \ldots, S_n$. In a (nonmultiple) wh-question, what the questioner is looking for is an individual, possibly depending on others. The *presupposition* of such a question (assuming it is in negation normal form) can be any sentence containing an existential quantifier or disjunction, say

(1) $S_0[(\exists x)S_1[x]]$.

An *answer* to the question is then of the form

(2) $S_0[S_1[f(y_1, y_2, \ldots, y_n)]]$

where $(\forall y_1), \ldots, (\forall y_n)$ are all the universal quantifiers within the scope of which $(\exists x)$ occurs in 1.

Of course, what in ordinary discourse is called an answer is often not a sentence like 2 but the substituting function or a substituting term.

What this means is that we interpret wh-questions as "for instance" questions. The questioner is not trying to identify all the individuals satisfying a certain condition. Coming to know one such individual will satisfy the inquirer. Likewise, in the propositional question whose presupposition is $(S_1 \vee S_2 \vee \ldots \vee S_n)$ one of the alternatives $S_1, S_2, \ldots, S_n$ is enough to serve the questioner's aims. Questions in which the purpose is to specify all the true or true-making alternations can be dealt with by the same conceptual tools as we are using, although we will not do so here.

It is to be noted that there also are questions where the alternatives are propositions, but propositions that depend on certain individuals. For such question can be taken to be of the form

$$S_0[(S_1 \vee \ldots \vee S_n)],$$

for instance

(3) $S_0[(S_1 \vee S_2)]$.

An answer (in our terminology) is then of the form

(4) $S_0[(S_1 \& f(y_1, \ldots, y_n) = 0) \vee (S_2 \& f(y_1, \ldots, y_n) \neq 0)]$

where $(\forall y_1), \ldots, (\forall y_n)$ are all the universal quantifiers within which $(S_1 \vee S_2)$ occurs in 3. When there are no such universal quantifiers, there are two possible answers, viz. $S_0[S_1]$ and $S_0[S_2]$.

The specification of the set Φ of the answers that an oracle will give is a part of the specification of a specific game of interrogative inquiry. It is assumed that Φ remains constant throughout an inquiry, and that an oracle will always answer a question when one of its answers is in Φ. It is also assumed that the new individual and function constant introduced by (L.E) and (R.A) do not occur in the members of Φ.

Together these rules define certain interrogative games (types of interrogative inquiry). However, strictly speaking they define only a cut-free version of interrogative inquiry. A more general type of interrogative inquiry is the defined by adding to rules so far introduced the following ones:

(L.taut) $$\frac{\Gamma \quad \rightarrow \quad \Delta}{\Gamma, (S \vee \quad \sim S) \rightarrow \quad \Delta}$$

(R.cont) $$\frac{\Gamma \quad \rightarrow \quad \Delta}{\Gamma \rightarrow \Delta \quad , (S \& \sim S)}.$$

Here S is an arbitrary formula of the underlying language. (L.taut) is to all practical purposes the same as proof theorists' "cut rule".

We will call an interrogative logic which includes (L.taut) and (R.cont) an *extended interrogative logic*.

All the rules for logical inferences (tableau building) and for questioning have been formulated here with an eye on a situation in which all answers are true and known to be true. We can call this case the logic of discovery. It turns out, however, that the same rules are basic for all interrogative logic. The main qualification need to this claim is that in a fully explicit interrogative logic we have to take into account also the epistemic element. (See Section 12 below.)

The interrogative derivability of C from the initial premises T in a model M will designated by

(5) $M : T \vdash C$.

This notion generalizes in an interesting way both the familiar notion of truth in a model, i.e. $M \vDash C$, and the equally familiar notion of logical

consequence, $T \vdash C$. The former relation is obtained as a special case of 5 when T is empty and all questions are answerable. The latter is obtained when no questions can be asked. These observations throw some light on the idea of truth as the ideal limit of inquiry.

4 SOME MAIN RESULTS

These rules, as they are so far formulated, capture only the simplest case of interrogative inquiry, viz. the case in which a single oracle's answers are all true (and known to be true by the inquirer). This might be considered as the case of pure discovery. The fact that we can develop an interesting theory of such discovery procedures is already highly suggestive, for among other things it shows conclusively the possibility of a genuine logic of discovery.

Such a logic is made interesting among other things by the possibility of establishing interesting metatheorems about the logic of interrogative inquiry. Some of them are extensions of the crucial metatheoretical results from logical theory to the logic of interrogative reasoning. Some are technically trivial, but interesting because of their applications.

THEOREM 1 (Completeness theorem). *If Φ is the set of all available answers and T the set of initial premises, then a conclusion C can be established in M by means of extended interrogative logic iff* $(\Phi \cup T) \vdash C$.

Proof. Let Φ be the set of all answers of the oracle.

Let us suppose that

$$M : T \vdash C.$$

In interrogative derivation of C from T in M we use logical moves and interrogative moves. If there are no interrogative moves then there is nothing to prove. Let $S_1, \ldots, S_n$ be the answers that we use in an interrogative derivation of C from T in M. So we have

$$T \cup \{S_1, \ldots, S_n\} \vdash C.$$

But by the definition of the set $\Phi, S_1, \ldots, S_n \varepsilon \Phi$. Hence we have

$$T \cup \Phi \vdash C.$$

This proves the "only if" part. Let us suppose coversely that

$$\Phi \cup T \vdash C.$$

By compactness of first-order logic there is a finite set $\Delta \subseteq \Phi$ such that

$$\Delta \cup T \vdash C.$$

Let $\Delta = \{S_1, \ldots, S_n\}$. Now the only problem is how we can have the sentence S_i as an answer to a question addressed to the oracle (for $i = 1, 2, \ldots, n$). Sentences $S_1, \ldots, S_n$ do not contain any dummy names that are introduced by the rules (L.Ef) or (R.Af). By using the rule (L.taut) we can introduce the tautologies $(S_i \vee \sim S_i)(i = 1, 2, \ldots, n)$ to the left side of subtableau. So the inquirer can ask the questions "S_i or $\sim S_i$?" $(i = 1, 2, \ldots, n)$. Now since $S_i \varepsilon \Delta \subseteq \Phi$ and all the sentences in the set Φ are true in the underlying model, we have $\sim S_i()\Phi$. Hence we have the sentence S_i as an answer to the corresponding question $(i = 1, 2, \ldots, n)$. Hence

$$M : T \vdash C.$$

■

The question of the completeness or incompleteness of unextended interrogative logic is a much messier one. It will not be discussed here.

Other results include the following:

THEOREM 2 (Yes-No Theorem). *In the extended interrogative logic, if $M : T \vdash C$, then the same conclusion C can be established by using only yes-no questions.*

A terminological explanation is in order here. For propositional question "Is it the case that S_1 or ... or S_n?" the presupposition is $(S_1 \vee \ldots \vee S_n)$. We say that a propositional question whose presupposition is of the form $(S \vee \sim S)$ is yes-no question.

Proof. Let us suppose that the answer to a propositional question "S_1 or ... or S_n?" is the sentence $S_i (i \leq n)$. Now instead of asking the propositional question, the inquirer can introduce the tautology $(S_i \vee \sim S_i)$ by using the rule (L.taut) and use it as the presupposition of the yes-no question "S_i or $\sim S_i$?". Clearly the sentence S_i is available answer to this question, for it was answerable as an answer to the given propositional question.

Let us suppose that the answer to the wh-question "$(\exists x)S[x]$?" is the sentence $S[b]$. Now instead of asking the wh-question, the inquirer can introduce the tautology $(S[b] \vee \sim S[b])$ by using the rule (L.taut) and use it as the presupposition of the yes-no question "$S[b]$ or $\sim S[b]$?". As in the case of any propositional question, the sentence $S[b]$ is available answer to that question.

The other cases can be dealt with similarly. ■

THEOREM 3 (Covering Law Theorem (special case)). *Let T be a theory in the first-order language L and P be a one-place predicate symbol occurring in T. Let $M : T \vdash P(b)$ where b is a given individual constant established by means of a set of answers A. Assume that A is consistent and that T does not entail $P(b)$. If P does not occur in A and if b does not occur in T, then there is a formula $H[x]$ such that*

1. $T \vdash (\forall x)(H[x] \supset P(x))$*;*

2. $A \vdash H[b]$*;*

3. Neither P *nor* b *occurs in* $H[x]$*.*

Proof. Let us suppose that $M : T \vdash P(b)$. Then obviously

$$A \cup T \vdash P(b).$$

By the compactness of first-order logic there is a finite subset T^* of T such that

$$A \cup T^* \vdash P(b).$$

Let $\wedge T^*$ be the conjunction of the members of the set T^*. Then we have

(6) $A \cup \{\wedge T^*\} \vdash P(b)$.

Now the deduction theorem of first-order logic gives us

(7) $A \vdash (\wedge T^* \supset P(b))$.

Here Craig's [1957a; 1957b] interpolation theorem can be seen to apply. Hence there exists an interpolant $I[b]$ such that

(8) $A \vdash I[b]$ and

(9) $I[b] \vdash (\wedge T* \supset P(b))$.

By the principles of first-order logic we obtain from 9

$$\wedge T^* \vdash (I[b] \supset P(b)).$$

Since b does not occur in T^*, we have

(10) $\wedge T^* \vdash (\forall x)(I[x] \supset P(x))$.

Hence we have

(11) $T \vdash (\forall x)(I[x] \supset P(x))$.

The vocabulary of $I[b]$ is shared by A and $(\wedge T^* \supset P(b))$. Thus P does not occur in $I[b]$ since it does not occur in A. Trivially b does not occur in $I[x]$. We get the intended result by putting $I[x] = H[x]$. ∎

Covering Law Theorem (suitably generalized) has important application in the philosophy of science, in particular in the theory of explanation. It can also be used to elucidate the nature of why-questions and of their role in interrogative inquiry. (See Section 19 below and [Hintikka and Halonen, 1995].)

THEOREM 4 (Covering Law Theorem (general version)). *Let T be a theory in the first-order language L and $F = F[b_1, \ldots, b_n]$ be a sentence such that $F[x_1, \ldots, x_n]$ is a formula of L and $b_1, \ldots, b_n \varepsilon do(M)$. Let $M : T \vdash F$ be established by means of a constant set of answers A. Let us assume that A is consistent and that T does not entail $F[b_1, \ldots, b_n]$. If other constants than $b_1, \ldots, b_n$ of F do not occur in A and if $b_1, \ldots, b_n$ do not occur in T, then there is a formula $H[x_1, \ldots, x_n]$ of L such that*

1. $T \vdash (\forall x_1) \ldots (\forall x_n)(H[x_1, \ldots, x_n]] \supset F[x_1, \ldots, x_n])$
2. $A \vdash H[a_1, \ldots, a_n]$
3. *none of the constants of* $F[x_1, \ldots, x_n]$ *occur in* $H[x_1, \ldots, x_n]$.

Proof. By assumption

$$A \cup T \vdash F[B_1, \ldots, b_n].$$

By the compactness of first-order logic there is a finite subset T^* of T such that

$$A \cup T^* \vdash F[b_1, \ldots, b_n].$$

In the same way as we obtained the formula 6 in special case we obtain here

(12) $A \vdash (\wedge T^* \supset F[b_1, \ldots, b_n])$

where $\wedge T^*$ is as in 6. By using Craig's interpolation theorem [1957a; 1957b] there exists an interpolan $I[b_1, \ldots, b_n]$ such that

(13) $A \vdash I[b_1, \ldots, b_n]$ and

(14) $I[b_1, \ldots, b_n] \vdash) \wedge T^* \supset F[b_1, \ldots, b_n])$.

Just as we obtained 11 we can here obtain

(15) $T \vdash (\forall x_1) \ldots (\forall x_n)(I[x_1, \ldots, x_n] \supset F[x_1, \ldots, x_n])$.

The vocabulary of $I[b_1, \ldots, b_n]$ is shared by A and $(\wedge T^* \supset F[b_1, \ldots, b_n])$. Thus, constants other than $b_1, \ldots, b_n$ of $F[b_1, \ldots, b_n]$ do not occur in $I[b_1, \ldots, b_n]$ since they do not occur in A. Constants $b_1, \ldots, b_n$ do not occur in $I[x_1, \ldots, x_n]$. We obtain the intended result by putting $H[x_1, \ldots, x_n] = I[x_1, \ldots, x_n]$. ■

The last result presented here will be called the *Strategy Theorem*. It is formulated only informally. (Cf. here [Hintikka, 1989].) It requires a preliminary explanation. If you review the questioning rule (L.Q), you can see that the formulas that can serve as presuppositions of questions are precisely the same formulas as the ones to which (L.E) or (L. $\vee f$) can be

applied. Hence there is a parallelism between the choice of one's interrogative strategy and the choice of one's deductive strategy. The question as to which formula to use as the presupposition of one's next question is paralleled by the question as to which formula to apply (L.E) or (L.$\vee f$) to next. The Strategy Theorem says that the two choices are governed by the same principles.

Above we have stated the rules which define the moves that may be carried out in deductive reasoning and in interrogative reasoning. These rules tell only how to reason (deductively and/or interrogatively) correctly. We will say that these rules are *definitory rules* for deduction and interrogation, respectively. These rules tell what is permitted in a game of reasoning. But we need also different kind of rules which define how to reason well, i.e. we need rules which tell how to use definitory rules effectively. We will call such rules *strategic rules* for deduction and interrogation, respectively. Such rules tell a reasoner how to argue effectively. Unfortunately there are no recursive rules which can be applied in every case. But it is possible to formulate some general strategic principles in deduction and interrogation. Among other things, we can establish a certain strategic parallelism between deduction and interrogation. (For an earlier discussion of the same idea, see [Hintikka, 1989].)

At each stage of an interrogative argument, only a finite number of applications of the rule (L.E) are possible. The sentences on the left side of the sequent reached at the time to which (L.E) can be applied are precisely the same ones that can serve as presuppositions of wh-questions. From studies of strategies of theorem-proving it is known that the crucial strategic question in deduction is which of these sentences to apply (L.E) to first. (Thus we can disregard the choice of propositional questions.) Obviously, in interrogative reasoning the crucial question is which of them to use first as the presupposition of a wh-question. With certain qualifications, it can be seen that not only is the range of choice the same in (L.E) and wh-questions, but also the crucial point; the optimal choice is the same in both cases.

This can be seen by examining the parallelism between deductive and interrogative steps. Suppose, for instance, that there occurs on the left side of a *(sub)tableau* an existential sentence $(\exists x)S[x]$. In deduction, the rule (L.E) says that the inquirer may add the sentence $S[\alpha]$ to the left side of a *(sub)tableau* where α is a new dummy name.

In interrogation the inquirer may use the rule (L.Q) and ask the wh-question "$(\exists x)S[x]$?". If the oracle answers then the inquirer may add the sentence $S[b]$ to the left column of the (sub)tableau where b is some individual from the domain of the underlying model.

Consider the difference between deduction and interrogation in the treatment of existential sentences. First, everything that can be done with $S[\alpha]$ can be done with $S[b]$, but not vice versa. The reason is that b is a real individual and α is only a dummy name. This means that we can ask

questions concerning the individual b but not concerning the dummy name α. To see what this means, let us suppose that there occurs on the left side of the same (sub)tableau an another existential sentence $(\exists x)S_1[x]$. In deduction the inquirer uses again the rule (L.e) and adds a new sentence $S_1[\beta]$ to the left side of the (sub)tableau where β is a new dummy name. In interrogation the inquirer may use information that he already has and use the rule (L.Q) to ask the yes-no question "$S_1[b]$ or $\sim S_1[b]$?". But in a similar way as above we see that if we use the rule (L.e) in deduction, then in interrogation we use the rule (Q.L) (if the question is answerable), and vice versa.

The general case $S_0[(\exists x)S[x]]$ can be treated similarly.

We can see that there are some differences between strategies in deduction and in interrogation. But we see that always when the inquirer in deduction uses in an optimal strategy one of the rules (L.E) or (L.$\vee f$), then the inquirer in interrogation should use the rule (L.Q) and ask the corresponding wh-question or propositional question, and vice versa.

Thus we have made plausible the following metatheoretical result:

THEOREM 5 (Strategy Theorem). *The optimal choice of the wh-question to ask is the same as the optimal choice of the rule (L.E) to apply, assuming that the question is answerable.*

In this form, the Strategy Theorem applies in the first place to unextended interrogative logic. It can be extended further by noting that (L.taut) and (R.cont) can be used in ordinary first-order logic to shorten proofs. In interrogative logic, they open the possibility of raising new yes-no questions and thereby facilitating new lines of reasoning. How closely parallel the principles of choosing the tautology for the two purposes remains to be studied.

The Strategy Theorem illustrates one of the major directions into which we are pushed in trying to develop a theory of interrogative reasoning. In doing so, we are soon led to realize the importance of strategies and strategic thinking, not only in the applications of interrogative logic, but in the very theory of this logic.

The strategic viewpoint also helps us to put into the proper perspective the Yes-No Theorem. It is virtually trivial technically and yet it seems to have sweeping implications. What are other kinds of questions good for, anyway, if their job could equally well be done by the simplest of all questions, yes-no questions?

The answer is that the purely inferential job of other kinds of questions can be done by yes-no questions, but that their strategic job cannot be so done. As one can see from the proof of the Yes-No Theorem, the yes-no question which does the same job as another one, say a simple wh-question, is found by means of the answer, say $S[b]$, to the wh-question. Clearly by asking "Is $S[b]$ the case or not?" the inquirer can obtain the

same statement as an answer to the a yes-no question. But how is the interrogative reasoner supposed to know to ask this particular question and not, say, a question about $S[d]$? Typically one does *not* know it without the help of earlier answer. Hence the "can" in the Yes-No Theorem is a definitory "can", not a strategic "can". It's import can perhaps be best expressed by saying that any interrogative argument can in the light of hindsight be subsequently replaced (in extended interrogative logic) by an argument where all questions used are yes-no questions.

5 A HIERARCHY OF LOGICS OF DISCOVERY

Even though our pure logic of discovery might at first sight appear exceedingly simple, it soon turns out to be a rich and intricate logical theory. The mere existence of such a logical theory amply shows that a logic of discovery in a perfectly natural sense is possible, even though the impossibility of such a logic has long been a dominant dogma in the philosophy of science. Indeed, it can be suggested that the logic of epistemological justification is more complicated and more difficult to have an overview of than the pure logic of discovery.

There is nevertheless no such thing as *the* logic of discovery. For one important thing, the structure of an interrogative process depends crucially on the class of answers Φ which the oracle(s) will give. (It is assumed to be constant throughout an interrogative inquiry.) How is Φ determined? In applications, there are usually various *ad hoc* restrictions on Φ which are hard to reach any overview on. For instance, in the case of scientific reasoning their study involves questions of observational and experimental methodology.

What can be studied here by means of logical methods is how our interrogative logic depends on the structural restrictions on available answers. The effects of such restrictions can be seen by examining the limiting case in which the structural restrictions are the only relevant ones. In this direction, the most important dimension is the quantificational complexity of the answers one's oracle (or oracles) can provide. (See [Hintikka, 1985].) The simplest case is the one in which all available answers are quantifier-free, i.e., particular propositions. In view of the Yes-No Theorem this case is to all practical purposes the one in which we restrict our attention to questions of the form "A or not-A?" where A is an atomic proposition. The assumption that we can restrict our attention in this way may be called the Atomistic Postulate. However, less narrow assumptions are likewise possible. The next case may be taken to be the one in which the available answers are either quantifier-free or of the form

$$(\forall x_1)(\forall x_2)\ldots S[x_1, x_2, \ldots, a_1, a_2, \ldots]$$

where S is quantifier-free. We will call this the A-case. The next one is the AE-case in which answers are also allowed to be of the form

$$(\forall x_1)(\forall x_2)\dots(\exists y_1)(\exists y_2)\dots S[x_1, x_2, \dots, y_1, y_2, \dots, a_1, a_2, \dots].$$

Sentences of this form are said to be AE sentences.

This hierarchy can obviously be continued. It is of a great importance for the logic of interrogative inquiry. Among other things, it enables us to characterize different types of inquiry from each other and likewise to classify different oracles according to the maximal logical complexity of the answers that they will give. For instance, purely observational inquiry is characterized by the Atomistic Postulate. In a diagnostic inquiry in medicine one oracle (the patient) provides only particular answers, whereas others (medical handbooks, the doctor's own background, biological knowledge, etc.) can in principle provide answers of any logical complexity.

6 THE ATOMISTIC CASE

The atomistic case results when the Atomistic Postulate is assumed. The result is a clear-cut and theoretically manageable logic of discovery. Let us assume that our interrogative logic is applied to a model ("world") M, and let us follow the usual practice of calling the set $D(M)$ of negated or unnegated atomic sentences true in M the *diagram* of M. (It is assumed that each member of the domain of individuals $do(M)$ of M has a name in the underlying language.) In the old-fashioned language of Carnap's $D(M)$ is essentially the state-description true in M. If no other restrictions are imposed on answers, we have $M : T \vdash C$ if and only if $\{D(M) \cup T\} \vdash C$.

In the atomistic case, the theory of interrogative logic thus reduces largely to ordinary logic, including its model theory. This enables us to establish a number of clear-cut results about the atomistic case. As an example, we can consider the question as to what an optimal theory looks like in interrogative logic. (Cf. here [Hintikka, 1984].) On the basis of ordinary deductive logic, an ideal theory T is of course a deductively complete theory, that is, a theory so that $T \vdash C$ or $T \vdash \sim C$ for every closed sentence C of the underlying language. Accordingly to what was just said, in interrogative logic the optimal theory T is such that

$$\begin{array}{ll} \{D(M) \cup T\} \vdash C & \text{or} \\ \{D(M) \cup T\} \vdash \sim C & \end{array}$$

for each C in the underlying language (including names for the members of the domain $do(M)$ of M) and for each model M of T. A moment's thought shows that such a theory is precisely what in model theory is known as a model-complete theory. (Cf. [Robinson, 1963].) Hence model-complete theories are the optimal ones if the Atomistic Postulate is applicable.

Most epistemologists and philosophers of science have tacitly assumed that the Atomistic Postulate is applicable in empirical reasoning. In such a case, the relevant oracle is nature. Admittedly, it looks plausible to assume that nature only tells us what happens on some particular occasion or other. She never gives general answers; she does not for instance tell what happens always and everywhere.

The Atomistic Postulate nevertheless soon leads into trouble or perhaps rather into a problem. For how can we then account in empirical terms for the kind of reasoning that leads to general theories? From particular propositions alone one cannot logically derive general theories. For the purpose, we also need strong initial premises T. But where do they come from? They cannot any longer be obtained empirically, for they are needed for the process of theory formulation. Hence what is needed is apparently some nondeductive principles of reasoning. These principles are usually referred to as principles of inductive inference. Thus one possible reaction to the problems caused by the adoption of the Atomistic Postulate is an inductivistic epistemology.

An inductivistic epistemology falls outside the scope of our discussion, however. Or, rather, the interrogative logic developed here is calculated among other things to replace inductivistic methodology, not to embody it. (Cf. below Sections 8 and 17.) We are not adopting the most popular alternative to inductivistic model of reasoning in epistemology, either, which is the hypothetic-deductive model of theoretical reasoning. According to this idea, the formation of general theories is not a rational process that can be dealt with any kind of "logic of discovery". It depends on intuition and serendipity rather than logic. Its results can be tested by comparing the deductive consequences of the resulting theory with experience, but the original discovery cannot be dealt with by means of a rational theory. In our view, such a view amounts to a failure to deal with one of the most important questions concerning human reasoning.

7 ON THE LOGIC OF EXPERIMENTAL REASONING

Our interrogative logic shows us a way out of this entire problem situation. The culprit here is not the problem of induction or the impossibility of the logic of discovery. It is the Atomistic Postulate. It simply is not true that the answers an inquirer receives from nature in an empirical inquiry are all particular propositions. The simplest counter-example is offered by controlled experiments. (They have in fact been considered ever since Kant as typical examples of nature's answers to an inquirer's questions to nature; see *Critique of Pure Reason* B xiii–xiv.) In a controlled experiment, a scientist is asking how the observed variable y changes when the controlled variable x is varied. An answer to such a question is an expression of

functional dependence which can be taken to have the logical form of a sentence of the AE form, e.g. of

$$(\forall x_1)(\forall x_2)\dots(\exists y_1)(\exists y_2)\dots S[x_1, x_2, \dots, y_1, y_2, \dots, a_1, a_2, \dots].$$

Hence the logic of experimental inquiry is an AE-logic rather than atomistic logic. It may even be argued that more complex propositions can be obtained as answers to experimental questions to nature. (Cf. here [Hintikka, 1988].)

The true logic of discovery therefore has at least the AE complexity. This does not mean that it cannot be studied in the same way as the logic of atomistic inquiry. The difference lies in the fact that instead of the diagram $D(M)$ of the relevant model we have to consider the set $D^{AE}(M)$ of all sentences of the AE form true in M — or perhaps some set which includes even more complex sentences.

There are results in effect concerning the interrogative AE logic in the literature, but it is not very easy to obtain a good overview on it. In any case, the AE logic is one of the most interesting and important forms which the logic of discovery can take. Its Protean character is an indication of its having been able to capture something of the richness of real-life nontrivial reasoning.

8 INTERROGATIVE INQUIRY BY MEANS OF UNCERTAIN ANSWERS

The basic interrogative logic so far developed is nevertheless too simple in several different ways. First, we must give up the assumption that all of the oracle's answers are true. This can be done surprisingly simply by rules that merely allow the inquirer to "bracket" any earlier answer (or initial premise) and conversely to "unbracket" a previously bracketed answer in the game *tableau*. The only catch here is that when an answer (or initial premise) is bracketed, all subsequent steps which depend on it must also be bracketed. This motivates the following rules which supplement the earlier ones:

(Q1) *Rule for Bracketing Lines:* At any stage of an interrogative game the inquirer may bracket any earlier answer or initial premise. We might for instance use a special kind of brackets will normally signal that the inquirer thinks the answer or premise is uncertain.

(Q2) *Rule for Bracketing Lines Derived from Bracketed Lines:* If a line in the table of an interrogative game is bracketed, all later lines dependent on it must also be bracketed.

(Q3) *Rule for Not Using Bracketed Lines in the Table:* When a line is bracketed, it may not be used in any application of the game rules, including the closure rules.

(Q4) *Rule for Unbracketing Lines:* At any stage of the game, the inquirer may unbracket any answer or premise bracketed at the time. Then all the later brackets caused by this line having been bracketed can also be removed.

The reason why we can get along with such simple rules is an important one. One might perhaps want us to formulate rules as to when an inquirer should bracket an answer. Such rules would nevertheless be strategic rules rather than what we can be called definitory ones. Such rules do not deal merely with what moves a player may make; they deal with the better and worse ways of applying the definitory rules of a game. It would be a serious mistake to try to incorporate such strategic rules into the definitory rules of one's game.

The detailed study of the strategic rules of bracketing and unbracketing nevertheless belongs to epistemology and philosophy of science rather than logic. In practice, such strategic rules of bracketing and otherwise evaluating an oracle's answers operate by means of evaluations of the reliability of the oracle in question. Conversely the basis of these evaluations is often the oracles previous track record as answering questions. As an extreme case, an oracle's answer can have two kinds of impact on the reasoner's thinking. It provides prima facie information whose reliability nevertheless has to be tested. But one such possible partial test is the answer itself, in that it can sometimes tell something about the oracle in question. For instance, it can show how well informed the oracle is and how well the new answer coheres with the rest of the inquirers relevant knowledge, which can provide some information about the oracle's reliability. Thus our interrogative logic provides us with an interesting vantage point which is likely to throw some light on the fascinating problem area of cognitive fallacies. (See [Piatelli-Palmarini, 1994] and the literature referred to there.) We will nevertheless not try to indulge in a full-scale discussion of these matters here. In this article, we will instead study more closely the logics of pure interrogative inquiry where all answers are true and known to be true.

9 INTERROGATIVE REASONING AND IDENTIFIABILITY

One of the many interesting applications of the ideas of interrogative logic concerns the notion of identifiability. (Cf. here [Hintikka, 1991].) Let T be a theory in a language L, and let P be a one-place predicate symbol occurring in T. We say that P is *explicitly definable* on the basis of T if and only if

(16) $T \vdash (\forall x)[P(x) \leftrightarrow D[x]]$

where $D(x)$ is a formula of L such that P does not occur in it and x is the only free variable in it.

It is possible to generalize the notion of definability by using our interrogative logic. We call this generalized notion *identifiability*.

We say that P is *explicitly identifiable* on the basis of T in a model M if and only if

$$M : T \vdash (\forall x)[P(x) \leftrightarrow D[x, a_1, \ldots, a_n]]$$

where $D[x, y_1, \ldots, y_n]$ is a formula of L such that P does not occur in it and $x, y_1, \ldots, y_n$ are the only free variables in $D[x, y_1, \ldots, y_n]$ and individuals $a_1, \ldots, a_n$ are from $do(M)(n < \omega)$. Auxiliary information may not contain P.

If there is no restriction on the available answers, then the explicit identifiability reduces to definability in the model M with parameters. If $n = 0$, and if the inquirer may not ask any questions then explicit identifiability reduces to explicit definability and if the inquirer uses questions in deduction then we have

(17) $M : T \vdash (\forall x)[P(x) \leftrightarrow D[x]]$

where $D[x]$ is as in an explicit definition of P. To get 17 we use auxiliary information from the model, so 17 is not equivalent to 16. Clearly, 16 entails 17, but not vice versa.

It is possible to define generalizations of other kinds of definitions, for example, we say that P is *piecewise identifiable* on the basis of T in a model M if and only if

$$M : T \vdash \bigvee_{i=1}^{i=k} (\forall x)[P(x) \leftrightarrow D_i[x, a_1, \ldots, a_n]]$$

where each $D_i (i = 1, \ldots, k)$ are as in the case of explicit identifiability.

Let Q be a new one-place predicate symbol. Let T' be the theory obtained from T by substituting Q for P in T. (The language $L(Q)$ used in T' is L extended by adding Q to it.) We say that P is *implicitly identifiable* on the basis of T in a model M if and only if

$$M : (T \cup T') \vdash (\forall x)[P(x) \leftrightarrow Q(x)].$$

Auxiliary information may not contain P or Q.

The following theorem is the very basic theorem in the theory of identifiability. (It is closely related to the Covering Law Theorem of Section 4.) Let C be a sentence of $L(Q)$.

THEOREM 6 (Extended Interpolation Theorem). *Assume that T is a consistent theory such that*

1. *$\mathbf{A} : T \vdash C$ and*

2. *not: $T \vdash C$.*

Then there is an interpolant $I[a_1, \ldots, a_n]$ such that

1. *each nonlogical constant of $I[a_1, \ldots, a_n]$ occurs in both T and C except for finite number of individual constants $a_1, \ldots, a_n$,*

2. *$\boldsymbol{A} : T \vdash I[a_1, \ldots, a_n]$,*

3. *$I[a_1, \ldots, a_n] \vdash C$.*

As a consequence of the Extended Interpolation Theorem we obtain the following Extended Beth's Theorem (cf. [Beth, 1953]):

THEOREM 7 (Extended Beth's Theorem). *P is explicitly identifiable on the basis of T in M if and only if P is implicitly identifiable on the basis of T in M.*

These theorems are proved in Hintikka and Harris [1988]. Extended Beth's Theorem helps us to use the techniques of interrogative logic in the theory of identifiability just as Beth's theorem helps us in the theory of definitions.

We get generalized notion of identifiability by saying that P is identifiable on the basis of T in every model of T. There are two different interpretations on this generalized notion: the defining formula $D[x, y_1, \ldots, y_n]$ is the same formula in every model of T and it can be different in different models. By means of this generalized notion of identifiability we can show some interesting interconnections between the notions of definability and identifiability. (See [Hintikka, 1991].)

The framework we have formulated enables us to make several useful general conceptual distinctions. For instance, we can now clear up some of the confusions that have affected the earlier discussion of the role of theoretical concepts in scientific theories. What needs to be done here is to distinguish the question of which terms can occur in the answers that the oracle (or the oracles) will provide from the question of which concepts occurring in a given theory T are identifiable. These are distinctly different questions. For instance, only the latter, but not the former, depends on the background theory. The former question might be called a question of observability, the latter a question of identifiability.

In practice, questions of observability seldom, if ever, have completely general answers. Even the most typical observable concepts, such as distance, are observable only over a certain range of values. Very large and very small distances have to be determined inferentially, instead of being directly observed. Hence it is questionable whether a sharp concept of observability is very useful in the first place.

In contrast, in questions of identifiability we can look away from the contingent restrictions on nature's answers and concentrate only on the logical structure of the theory T plus the structural restrictions on nature's answers. It is likely to play an important role in the general methodology of science, as it has already played in the methodology of some particular sciences, for instance in econometrics. (Cf. here [Hsiao, 1983].) For one thing, classical problems of definability in science frequently turn out to pertain to identifiability rather than definability in the strict sense of the word used here.

10 SEMANTICAL GAMES AND INFORMATIONAL INDEPENDENCE

The simple logic of interrogation so far outlined is in many ways instructive as an approximation to the full explicit logic of questioning. Nevertheless it must be refined in several ways. Perhaps the most important weakness it has is that it can only be used to study interrogative arguments with a predetermined ultimate conclusion. This does not deprive it of all interest. It can be used to study the confirmation and explanation of conclusions. It can be used for the purposes of answering propositional questions by means of interrogative inquiry. Suppose you want to answer by means of interrogative inquiry the question whose presupposition is

$$(S_1 \vee S_2 \vee \ldots \vee S_n)$$

starting from the initial premises T. All you need to do is to try to construct n parallel *tableaux* which start from the sequents

$$T \rightarrow S_i$$

$(i = 1, \ldots, n)$. This treatment applies *a fortiori* to yes-no questions, but it cannot be extended to questions where a disjunction is independent of K but dependent on one or more universal quantifiers. Most importantly, the simple form of interrogative inquiry so far studied cannot be used for the purpose of answering wh-questions by means of interrogative reasoning. Yet such processes of answering "big" (principal) questions by putting a number of "small" (operative) questions to one's oracles is the most common and the most important use of interrogative reasoning. This double role of questions is instantiated in Plato's dialogues where Socrates examines an overarching question ("What is knowledge?", "What is piety?" etc.) by putting a number of smaller questions to his interlocutor. Likewise, Aristotle was aware of the distinction between principal and operative questions and warned us against confusing the two.

Furthermore, the simple form of our interrogative logic does not explain the relationship between questions, their presuppositions, and their (conclusive) answers. For not any singular term of the appropriate sort can qualify as an answer in the sense of a conclusive answer, that is to say, as an answer that does the job it is semantically speaking calculated to do. For instance, if I ask,

(18) Who murdered Roger Ackroyd? or

(19) Who is Lord Avon?

you are not providing a conclusive answer if you respond by saying

(20) Whoever murdered him or

(21) Lord Avon, who else?

Both these difficulties can be overcome in the same way. What we have to do is to construe interrogative inquiry as a search of knowledge and not just as an exercise in argumentative bridge-building. This means enrolling epistemic logic to our service. The initial premises are now expressed by the inquirer by "I know that S_i", where S_i is one of the propositions that previously was called an initial premise. They will be abbreviated by $K_I S_i$. Since only one inquirer is considered in this paper, we will omit in the sequel the subscript to K which indicates the knower. An answer to a question will also have the form KA, as will the ultimate conclusion.

In order to develop a viable epistemic logic that could help us here, we first have to formulate a semantics for epistemic statements. This is best done (for reasons which soon will become clear) in game-theoretic form. In the simplest case (which is the only one dealt with explicitly here) we are dealing only with the case of a single knower. Hence we can simply use K without a subscript. Then the interpretation of an epistemic sentence is with respect to a model system which is a set Ω of models of the (nonepistemic part of the) underlying language, together with a two-place relation defined on this set. This relation is called the epistemic alternativeness relation. Intuitively, given a model ("possible world") M, the epistemic alternatives to M are all the worlds that are compatible with everything the knower in question knows in M.

Then the rules for semantical games are obvious. Such a game begins in the model in which the given sentence S is evaluated. At each stage, the players are considering a sentence and a model.

(R.&) $G((S_1 \& S_2))$ begins with the falsifier's choice of S_i ($i = 1$ or 2). The game is then continued as in $G(S_i)$.

(R. $\vee$) $G((S_1 \vee S_2))$ begins with the verifier's choice of S_i ($i = 1$ or 2). The game is then continued as in $G(S_i)$.

(R.A) $G((\forall x)S[x])$ begins with the choice by the falsifier of a member, say b, of $do(M)$. (If this individual does not have a name in the underlying language L, it is given one which is added to L.) The rest of the game is as in $G(S[b])$, where $S[b]$ is the sentence obtained from $(\forall x)S[x]$ by omitting the quantifier and by replacing all occurrences of the variable bound to it by "b".

(R.E) Likewise, except that choice is made by the verifier.

(R.$\sim$) $G(\sim S)$ begins with an exchange of roles of the two players, as defined by the game rules (including rules for winning and losing). The rest of the game is otherwise as in $G(S)$.

Semantical games of this sort must be distinguished sharply from the epistemic questioning games. The distinction and its significance is spelled out in [Hintikka, 1996b, Chapter 2] and [Hintikka, 1995]).

11 A NEW EPISTEMIC LOGIC FOR DIFFERENT TYPES OF KNOWLEDGE

The game-theoretical nature of the semantics we have developed enables us to take the crucial step needed to develop a viable logic of knowledge, questions and answers. It is to allow the moves connected with disjunctions and existential quantifiers (when made by the initial verifier, ditto for moves for conjunctions and universal quantifiers, when likewise made by the initial verifier) informationally independent of a sentence-initial K. (Cf. [Hintikka, 1996b, Chs 4 and 11].) Such independence provides the precise logico-semantical realization of the question ingredient present in natural languages. We will indicate such independence by a slash notation, writing e.g. $(\exists x/K)$ instead of $(\exists x)$ and $(\vee/K)$ instead of $\vee$.

By means of this notation we can express the logical forms of all the different types of knowledge.

Here we focus (for reasons of space) on what might be called nonmultiple (single-target) knowledge, excluding such kinds of knowledge as involve the specification of two different items. For instance, sentences like the following will not be directly treated here:

Mary knows to whom John gave what

even though it can be represented (on its different readings) in our treatment as

$$K_{\text{Mary}}(\exists x/K_{\text{Mary}})(\exists y/K_{\text{Mary}})(\text{John gave } y \text{ to } x)$$

or as

$$K_{\text{Mary}}(\forall y)((\exists x)(\text{John gave } y \text{ to } x) \supset (\exists x/K_{\text{Mary}})(\text{John gave } y \text{ to } x)).$$

Indeed our discussion of nonmultiple questions can be without any problems extended to multiple questions. Their theory has been studied (in a framework not involving informational independence) in [Hintikka, 1976].

The kinds of knowledge considered here are represented by *knowledge statements*. They are of the form KS where S is an ordinary first-order statement (in negation normal form) except that one of the existential quantifiers $(\exists x)$ has been replaced by $(\exists x/K)$ or one of the disjunctions $(S_1 \vee S_2)$ has been replaced by $(S_1(\vee/K)S_2)$. Also, K_I is shortened to K. Our treatment of this case can be extended to knowledge statements containing several slashes.

The following list of knowledge statements and their formalizations illustrates the power and flexibility of our "independence-friendly" slash notation.

(22a) I know that S

(22b) KS

(23a) I know whether S

(23b) $K(S(\vee/K) \sim S)$

(24a) I know whether S_1 or S_2

(24b) $K(S_2(\vee/K)S_2)$

(25a) I know who (say x) is such that $S[x]$

(25b) $K(\exists x/K)S[x]$

(26a) I know to whom each person bears the relation $S[x, y]$

(26b) $K(\forall x)(\exists y/K)S[x, y]$

(27a) I know of each person (say x) whether he or she satisfies the condition $S_1[x]$ or $S_2[x]$

(27b) $K(\forall x)(S_1[x](\vee/K)S_2[x])$.

Of these, (22)–(25) have an equivalent slash-free first-order representation, while (26)–(27) do not.

These different kinds of constructions illustrate the role of informational independence in knowledge statements. Consider, for instance, a statement of the form

(28) $K(\exists x/K)M(x, r)$

as distinguished from

(29) $K(\exists x)M(x,r)$

where $M(x,r)$ might for instance be "x murdered Roger Ackroyd". Then (28) says merely that I know that someone murdered Roger Ackroyd. In contrast, (29) says that in each world compatible with my knowledge there is an individual *chosen independently of the world in question* and hence the same in all of them. In other words, there is an individual such that my knowledge rules out all possibilities that individual did not murder Roger Ackroyd. And this obviously means that *I know who* murdered Roger Ackroyd.

In the case of (28) the same effect could be obtained by operator ordering in that (28) is logically equivalent with $(\exists x)KM(x,r)$. But an existential quantifier, even though independent of K, may nevertheless depend on some intermediate universal quantifier, as in

(30) $K(\forall x)(\exists y/K)S[x,y]$.

Then the knowledge statement does not have any slash-free equivalent

12 AN EPISTEMIC LOGIC FOR QUESTIONS AND ANSWERS

These different kinds of knowledge statements (with the exception of simple *knows that* statements) can operate as *desiderata* of so many kinds of questions. The desideratum of a question specifies the epistemic state that a conclusive answer to it calculated to bring about (on the assumption that the presupposition of the question is true) in a typical information-acquiring use of questions. The direct questions whose desiderata (23)–(26) are can accordingly be expressed as follows:

(31c) Is it the case that S?

(32c) Is it the case that S_1 or S_2?

(33c) Who (say x) is such that $S[x]$?

(34c) To whom does each person bear the relation $S[x,y]$?

(35c) Does each person (say x) satisfy the condition $S_1[x]$ or $S_2[x]$?

Like (22a)–(27a), (31c)–(35c) are not idiomatic English sentences. However, it is easy to construct such sentences with the same logical form as (22a)–(17a) and (31c)–(35c).

Our slash notation enables us formulate all the basic concepts concerning questions and answers. For instance, the *presupposition* of a question can be obtained from its desideratum simply by omitting all the slashes, for instance replacing $(\exists x/K)$ by $(\exists x)$. Thus the presuppositions of (31c)–(35c) are:

(31d) $K(S \vee \sim S)$

(32d) $K(S_1 \vee S_2)$

(33d) $K(\exists x)S[x]$

(34d) $K(\forall x)(\exists y)S[x, y]$

(35d) $K(\forall x)(S_1[x] \vee S_2[x])$.

A *reply* results from the desideratum by omitting the slashed quantifiers and replacing the variables bound to one of them by an individual or function constant of the appropriate kind. For instance, the following are the logical forms of replies to (31c)–(35c):

(31e) $KS, K \sim S$

(32e) KS_1, KS_2

(33e) $KS[b]$

(34e) $K(\forall x)S[x, f(x)]$

(35e) $K(\forall x)((S_1[x] \& f(x) = 0) \vee (S_2[x] \& f(x) \neq 0))$.

A reply is a *conclusive answer* if and only if the identity of the referents of all the substituting individual and function constants is known. For instance, the conclusiveness conditions for (33e)–(35e) can be expressed as follows:

(33f) $K(\exists x/K)(b = x)$

(34f) $=$ (32f) $K(\forall x)(\exists y/K)(f(x) = y)$.

This is logically equivalent with either of the following

$$K(\exists g/K)(\forall x)(f(x) = g(x))$$
$$(\exists g)K(\forall x)(f(x) = g(x))$$

which show the parallelism between (33f) and (34f).

It can be shown that a reply to a question logically implies its desideratum if the corresponding conclusiveness condition is satisfied. This shows that the conclusiveness conditions really do the job they were calculated to do, which is to characterize the relationship of a question to its (conclusive) answers. The weak point in all the earlier approaches to the logic and semantics of questions and answers (cf., e.g. [Collingwood, 1933; Belnap and Steel, 1976; Harrah, 1984; Bromberger, 1992; Wisniewski, 1995]) is that this fundamental relationship between questions and their (conclusive) answers is not done justice in them.

Replies to a given question which do not satisfy the conclusiveness condition can convey to the questioner some useful information, albeit not all the information needed to satisfy the questioner. Such replies can be called *partial answers*. They deserve (and need) a separate study.

13 CONCLUSIVENESS CONDITIONS AS CODIFYING CONCEPTUAL KNOWLEDGE

Thus we have reached a deeper understanding of the nature of answers to questions in the interrogative inquiry — and of the nature of questions and answers in general. One striking fact here is that two different things are required for a conclusive answer to a question, on the one hand a reply to it and on the other hand the conclusiveness condition for that reply. In general, these two do not always come from same source or represent the same kind of information. For instance, consider an experimental question whose desideratum is of the form

(36) $K(\forall x)(\exists y/K)S[x,y]$.

It is a question concerning how the variable y depends on the variable x in $S[x,y]$. A reply to such a question is of the form

(37) $K(\forall x)S[x,g(x)]$.

Here we can think of the function $g(x)$ as being given extensionally as a class of ordered pairs of values of the controlled variable x and corresponding measured values of the observed variable $y = g(x)$. Such a way of being given does not yet tell the inquirer which mathematical function $g(x)$ is. This additional information is what is expressed by the conclusiveness condition

(38) $K(\forall x)(\exists y/K)(g(x) = y)$

which may also be expressed as

(39) $K(\exists f)(\forall x)(g(x) = f(x))$

or even

(40) $K(\exists f)(g = f)$.

This shows that what the conclusiveness condition (38) says, is that the inquirer knows which function the function $g(x)$ is that figures in the given reply, i.e. knows the (mathematical) identity of this function.

But if so, it is clear that the information (38) yields to the inquirer is not given to him or her through the experiment itself. It constitutes *a priori* mathematical knowledge which comes to him from an entirely different

source. As we might illustrate the difference, in order to obtain the reply (37) a theoretical physicist has to appeal to his experimentalist colleagues, but in order to reach the conclusiveness condition (38) he or she has to sit with mathematicians and commit *a priori* reasoning.

Hence it is reasonable to require of an interrogative step only that a reply to the question in question is given by the oracle. What is needed then is another oracle that answers questions of identification, i.e. questions concerning which objects of different logical types can be considered as being known by the inquirer. We might call this the identification oracle. The need to postulate such an oracle as a part of the interrogative model shows that *a priori* truth does not enter empirical inquiry only through the logical inference steps of inquiry. An independent source of such truths is in principle intrinsic to all reasoning and inquiry. We will call such a source of answers an *a priori* oracle in contradistinction to ordinary *a posteriori* ones, and we will apply the same terms to the steps of consulting the one or the other kind of oracle.

However, a further explanation is in order here. In one respect, the term *a priori* knowledge is misleading. If we look at the role of conclusiveness conditions in interrogative inquiry, we can see that they have to be reached in principle in the same way as all the other conclusions which serve as stepping-stones to further ones, viz. by means of the same interrogative logic. They need not be given, as initial premises or in any other way. In this respect, it may very well be misleading to call them *a priori*. In general, there is no reason not to think of mathematical and other conceptual knowledge as also being reached by means of interrogative reasoning, not unlike ordinary empirical one. However, this does not eliminate a distinction between the two in that they rely on the testimony of different kinds of oracles.

One further comment is needed here. Usually it does not make sense to say that the choices of the values of two adjacent universal quantifiers are dependent or independent of each other. Hence there seems to be no difference between $K(\forall x)$ and $(\forall x)K$, for K is in effect a universal quantifier ranging over epistemic alternatives. By the same token, the distinction between $K(\forall x)$ and $K(\forall x/K)$ might seem to be without difference. But a second thought shows that the ranges of universal quantifiers inside and outside the scope of K are different. In $K(\forall x)$ the values are the individuals existing in the world whose choice is governed by K, whereas in $(\forall x)K$ and $K(\forall x/K)$ the choice is from the individuals that are well defined for all the epistemic alternatives. Hence we need also slashed quantifiers of the form $(\forall x/K)$.

14 THE TWO ROLES OF QUESTIONS IN INTERROGATIVE LOGIC

By means of the epistemic logic and the theory of questions and answers so developed, we can study the logic of interrogative inquiry in greater detail than before. We will call the result epistemic interrogative logic. The first main change is that the initial premises and the ultimate conclusion (the statement to be established) are knowledge statements. Our task is then to reformulate the rules of interrogative inquiry which now involve an epistemic element and then study the resulting epistemic interrogative logic.

The principal novelty here is the possibility of using the desideratum of a question (i.e. of the principal question of an inquiry) as the ultimate conclusion. Then the interrogative argument leading to that conclusion (assuming that the game *tableau* is closed) is tantamount to answering the principal question by means of questioning.

In other words in such an interrogative inquiry the principal question is answered with the help of "small" or operational questions put to the oracle. This is eminently in keeping with the historical precedents of our interrogative logic. The main difference is that now we have as our disposal an explicit theory of questions, answers and the question-answer relationship.

In this way we can bring our interrogative logic to bear on problem-solving. In some ways, the extension (epistemic version) of our interrogative logic can be thought of as a general logic of problem-solving. This approach is especially natural in that our treatment will automatically involve a treatment of the different strategies of problem-solving. In such a treatment, the Strategy Theorem of Section 4 above (together with its extensions) is obviously going to play a crucial role.

The other main advantage offered by an epistemic approach to interrogative inquiry is a closer analysis of the nature of the answers that are used in interrogative inquiry as new premises. To take an example, in the first-order approximation to a logic of interrogative inquiry, an answer to a question whose presupposition is

(41) $(\forall x)(\exists y)S[x,y]$

was of the form

(42) $(\forall x)S[x,g(x)]$.

It does not seem to matter very much to the logical relations of these two statements whether we prefix both of them by K or not. But in a deeper perspective we cannot restrict our attention any longer to (41)–(42) or to their epistemic versions

(43) $K(\forall x)(\exists y)S[x,y]$

(44) $K(\forall x)S[x,g(x)]$.

For we know now, on the basis of the analysis of questions and answers offered in Section 12 above, that the function $g(x)$ does not automatically yield a conclusive answer, i.e., that (44) is not always a conclusive answer to the question to the question whose presupposition is (43). What is also needed is the conclusiveness condition

(45) $K(\forall x)(\exists y/K)(g(x) = y)$.

This conclusiveness condition could also be written in the following forms:

(46) $K(\exists f/K)(\forall x)(g(x) = f(x))$

(47) $(\exists f)K(\forall x)(g(x) = f(x))$

(48) $(\exists f)K(g = f)$.

This shows the intuitive import of the conclusiveness condition in this case. What (44) says is that the function $g(x)$ is of the kind desired as an answer. What (48) says, i.e., what the conclusiveness condition says is that it is known which function $g(x)$ is.

Hence in the nonepistemic (first-order) formulation of the logic of interrogative inquiry we had to assume of each available answer that the relevant conclusiveness condition is satisfied. This impoverishes one's interrogative logic greatly. For one thing, the reply (4) and the conclusiveness condition (45) might come from different sources. For instance, the function $g(x)$ might be given (in extension, so to speak) by a scientific experiment. Such a function would not yield a satisfactory answer unless or until it is known which mathematical function we are dealing with. And the information that makes that condition true is likelier to come from mathematics than from a physical experiment.

Thus the epistemic approach opens up the possibility of a sharpen analysis of the role of answers and of their conclusiveness conditions in interrogative inquiry. Indeed, our comments on (41)–(48) concern essentially what was called above in section 5 the AE-case, which includes the logic of experimental inquiry. A moment's thought shows that without taking the epistemic element into account we cannot even do justice to the presuppositions of such questions, let alone to the conclusiveness conditions of their answers.

15 EPISTEMIC RULES FOR INTERROGATIVE LOGIC

But what do the new, improved rules for interrogative inquiry look like? At first sight it might seem impossible to formulate a complete set of rules for the purpose. This impossibility is suggested by independence-friendly first-order logic, for which a complete set of rules of proof cannot be formulated,

even though there is a complete set of rules of disproof, that is, for proving logical inconsistency. This is connected with the fact that IF first-order languages are not closed with respect to contradictory negation. (See here [Hintikka, 1996b, Chs. 3–4, 6].) Hence we cannot show that $(F \supset G)$ is valid by showing that there is a model in which F is true and G not true, for G's not being true cannot be expressed by any IF first-order formula. The simplest sentences whose contradictory negations fail to be expressible in an IF first-order language are the Henkin quantifier formulas, i.e., formulas of the form

(49) $(\forall t)(\exists x)(\forall y)(\exists z/\forall t)S[t, x, y, z]$.

Now knowledge statements may seem to exhibit this same quantifier structure when we realize that the knowledge operator K is a disguised universal quantifier. For instance, knowledge statements of the following form may seem to have the same difficulties associated with them as Henkin quantifier sentences the IF first-order logic:

(50) $K(\exists x)(\forall y)(\exists z/K)S[x, y, z]$.

Luckily this first impression turns out to be incorrect. The reason is that the "quantifier" K is in a certain sense nonstandard. In order to see this subtle point, consider a knowledge statement of the form

(51) $K(\forall x)(\exists y/K)S[x, y]$.

This statement is true if and only if

(52) $(\exists f)K(\forall x)S[x, f(x)]$.

But if f is considered as a function of two variables, x and k, the relevant epistemic alternative $f(x, k)$ cannot be chosen arbitrarily. It has to be a function of the *individual* x which is well defined for all the epistemic alternatives. It follows that whenever the value of x in two alternatives is the same, the value of $f(x)$ in these two worlds, must also coincide. This restricts the set of the functions f over which the function quantifier $(\exists f)$ ranges and turns this quantifier in effect into a first-order quantifier ranging over the world lines of individuals well-defined in all alternatives. For this reason we may hope to be able to formulate a complete set of rules for interrogative inquiry, in spite of the presence of the phenomenon of informational independence.

In formulating the new rules for interrogative logic we will assume throughout that the formulas we are dealing with are in the negation normal form.

The rules for interrogative logic involving epistemic notions can be formulated in terms of what we will call k-sequents. We can think of them as ordinary sequents in which each formula is thought of as containing a suppressed initial K and in which there is initially only one formula on the right side. In terms of such k-sequents, we can formulate the following rules:

(L.&), (L.∨f), (R.& f), (R. ∨), (L.E), (R.A), (L.A), (R.E) as before.

$$\text{(L.E/K)}\quad \frac{\Gamma, S_0[(\exists x/K)S_1[x]] \rightarrow \Delta}{\begin{array}{l}\Gamma, S_0[S_1[f(y_1,\ldots,y_i)]],\\ (\forall y_1)\ldots(\forall y_i)(\exists x/K)(f(y_1,\ldots,y_i)=x) \rightarrow \Delta.\end{array}}\,.$$

Here f is a new function symbol and $(\forall y_1),\ldots,(\forall y_i)$ are all the universal quantifiers within the scope of which $(\exists x/K)$ occurs in $S_0[(\exists x/K)S_1[x]]$.

$$\text{(R.E/K)}\quad \frac{\Gamma, (\exists x/K)(t=x) \rightarrow \Delta, S_0[(\exists xK)S_1[x]]}{\Gamma, (\exists x/K)(t=x) \rightarrow \Delta, S_0[(\exists x/K)S_1[x]], S_0[S_1[t]]}$$

where t is any constant term.

In our epistemic version of interrogative logic, the counterpart to the tautology introduction rule will be the following

$$\text{(L.taut)}\quad \frac{\Gamma, \Delta}{\Gamma \rightarrow K(S \vee \sim S) \rightarrow \Delta}$$

where S is any formula in the underlying language.

The rule (R.cont) cannot be formulated quite analogously, because there must not be more than one formula on the right-hand side of the sequent:

$$\frac{\Delta \rightarrow KS_1}{\Gamma \rightarrow K(S_1 \vee (S_2 \& \sim S_2))}$$

where S_2 is any formula in the appropriate language.

We need also rules for independent disjunctions.

$$\text{(L. }\vee/K)\quad \frac{\Gamma, S_0[S_1(\vee/K)S_2] \rightarrow \Delta}{\begin{array}{l}\Gamma, S_0[(S_1 \& f(y_1,\ldots,y_i)=0) \vee (S_2 \& f(y_1,\ldots,y_i) \neq 0)]\&\\ (\forall y_1)\ldots(\forall y_i)(\exists x/K)(f(y_1,\ldots,y_i)=x) \rightarrow \Delta\end{array}}$$

where f is a new function symbol and $(\forall y_1),\ldots,(\forall y_i)$ are all the universal quantifiers within the scope of which $(S_1(\vee/K)S_2)$ occurs in S_0.

$$\text{(R. }\vee/K)\quad \frac{\Gamma \rightarrow \Delta, S_0[S_1(\vee/K)S_2]}{\Gamma \rightarrow \Delta, S_0[(\exists x/K)((S_1 \& x=0) \vee (S_2 \& x \neq 0))]}\,.$$

16 ON THE ROLE OF CONCLUSIVENESS CONDITIONS

The way these rules operate can be illustrated by showing how a reply and the conclusiveness condition in some selected cases imply the desideratum of a question. In the following proofs, we are leaving all appeals to structural rules (i.e. rules for permuting members of sequents) tacit. The sequents involved are of course k-sequents.

(i) $(\forall x)S[x,g(x)], (\forall x)(\exists y/K)(g(x)=y) \to (\forall x)(\exists y/K)S[x,y]$
(Cf. here (43)–(45).)

(ii) $(\forall x)S[x,g(x)], (\forall x)(\exists y/K)(g(x)=y) \to (\exists y/K)S[b,y]$
from (i) by (R.A).

(iii) $(\forall x)S[x,g(x)], (\forall x)(\exists y/K)(g(x)=y), S[b,g(b)], (\exists y/K)(g(b)=y) \to (\exists y/K)S[b,y]$
from (ii) by two application of (L.A) (and of structural rules).

(iv) $(\forall x)S[x,g(x)], (\forall x)(\exists y/K)(g(x)=y), S[b,g(b)], (g(b)=d),$
$(\exists y/K)(d=y) \to (\exists y/K)S[b,y]$
from (iii) by (L.E/K).

(v) $(\forall x)S[x,g(x)], (\forall x)(\exists y/K)(g(x)=y), S[b,g(b)], g(b)=d,$
$(\exists y/K)(d=y) \to (\exists/K)S[b,y], S[b,d]$
from (iv) by (R.E/K).

(vi) $(\forall x)S[x,g(x)], (\forall y)(\exists y/K)(g(x)=y), S[b,g(b)], g(b)=d,$
$(\exists y/K)(d=y), S[b,d] \to (\exists y/K)S[b,y], S[b,d]$
from (v) by (L.=). Here (vi) is a closure sequent.

Likewise the following k-sequent can be proved.

(i) $(\exists x)(\forall y)S[x,y,g(y)], (\forall x)(\exists y/K)(g(x)=y)$
$\to (\exists x)(\forall y)(\exists z/K)S[x,y,z]$.
The proof might run as follows:

(ii) $(\forall y)S[a,y,g(y)], (\forall x)(\exists y/K)(g(x)=y) \to (\exists x)(\forall y)(\exists z/K)S[x,y,z]$
from (i) by (L.E) (and structural rules).

(iii) $(\forall y)S[a,y,g(y)], (\forall x)(\exists y/K)(g(x)=y)$
$\to (\exists x)(\forall y)(\exists z/K)S[x,y,z], (\forall y)(\exists z/K)S[a,y,z]$
from (ii) by (R.E).

(iv) $(\forall y)S[a,y,g(y)], (\forall x)(\exists y/K)(g(x)=y)$
$\to (\exists x)(\forall y)(\exists z/K)S[x,y,z], (\exists z/K)S[a,b,z]$
from (iii) by (R.A).

(v) $(\forall y)S[a,y,g(y)], (\forall x)(\exists y/K)(g(x)=y), S[a,b,g(b)], (\exists y/K)(g(b)=y)$
$\to (\exists x)(\forall y)(\exists z/K)S[x,y,z], (\exists z/K)S[a,b,z]$
from (iv) by (L.A) (and of structural rules).

(vi) $(\forall y)S[a,y,g(y)], (\forall x)(\exists y/K)(g(x)=y), S[a,b,g(b)], (\exists y/K)(g(b)=y)$
$\to (\exists x)(\forall y)(\exists z/K)S[x,y,z], (\exists z/K)S[a,b,z], S[a,b,g(b)]$
from (v) by (R.E/K). Here (vi) closes the tableau construction and thereby shows that (i) is valid.

Thus an implication from a reply and its conclusiveness condition to the desideratum of a question holds also when the conclusion prima facie has a Henkin quantifier form. This is an indication of the fact mentioned earlier, viz. that the presence of Henkin-type prefixes (as in (50) above) in epistemic logic does not prevent the formulation of a complete sets of rules for interrogative logic.

17 REPLIES AND ANSWERS IN INTERROGATIVE REASONING

The question-answer relationship codified by the conclusiveness condition is the crucial novelty of the approach to the logic of questions and answers ("erotetic logic") presented here. Most of the other approaches are reduced to labeling every response to a question an answer which provides the questioner with some relevant information. It is easy to find a wealth of examples of such responses which nevertheless do not satisfy the questioner but instead prompt further questions.

It is nevertheless important to be clear precisely what role the requirement of conclusiveness plays in the questions and answers that figure in interrogative reasoning. In Section 14 it was pointed out as a weakness of the simplified nonepistemic version of interrogative logic (Sections 1–9 above) that oracles' responses must all be assumed to be conclusive answers. A stipulation to this effect was implicitly made in the original question rule. This would nevertheless be unrealistic in many applications of our interrogative logic. Now we are in a position to see how the question rule can be made more realistic. What often, perhaps typically, happens in ordinary discourse is that a reply and the corresponding conclusiveness condition are established independently of each other. Only when they are put together do they imply a conclusive answer to a question. We must clearly allow for such a procedure also in our interrogative logic. That means that in the question rule an oracle's response can be a reply and not always a conclusive answer. We can — and perhaps should — require that if an oracle can give a conclusive answer, it will do so. But when there is no conclusive answer to a given question among the ones an oracle can yield, we can allow the oracle to respond by giving a reply to it in our technical sense. The crucially important role of conclusiveness condition in a systematic theory of questions, answers, and interrogative logic does not mean that an oracle's responses always have to satisfy them. Some requirements of informativeness must nevertheless be imposed on an oracle's replies, in order to avoid the possibility that an obstreperous oracle can frustrate an interrogative argument altogether. We might even try to impose a requirement of maximal informativeness on oracles' replies. In particular, it is in keeping with the spirit of most applications to require an oracle who is answering basic questions of identification to give maximally informative replies. These are questions

whose desideratum is of the form

$$K(\forall y_1)\dots(\forall y_i)(\exists x/K)(f(y_1,\dots,y_i)=x)$$

if the oracle can give a conclusive answer, it will give one as a reply.

What these modifications, the question rule allows more realistic applications of our interrogative logic than before. Among other things, as was noted, our new rule makes it possible that the inquirer should receive a reply to a question and the corresponding conclusiveness condition separately, perhaps as replies to different questions. This is realistic. In many situations, establishing the conclusiveness condition of a reply is the responsibility of the inquirer, not of the answerer. For instance, it is unrealistic to expect nature to yield more than replies to questions. The inquirer has to gather the corresponding conclusiveness condition from other sources.

18 SOME MAIN RESULTS OF EPISTEMIC INTERROGATIVE LOGIC

For epistemic interrogative logic, we can establish counterparts to the four basic metatheorems formulated in Section 4 above.

THEOREM 8 (Completeness Theorem). *If a knowledge statement KS is logically implied by* Γ *plus the set of all answers available, then*

$$\Gamma \rightarrow KS$$

can be proved in the extended epistemic interrogative logic.

THEOREM 9 (Yes-No Theorem). *In extended epistemic interrogative logic, if a sequent*

$$\Gamma \rightarrow KS$$

can be established interrogatively, then it can also be established by using only yes-no questions.

These two metatheoretical results are fairly straightforward extensions of the earlier ones. Their proofs do not prompt any surprises.

The Covering Law Theorem for the nonmodal interrogative logic relived on Craig's interpolation theorem for first-order logic. In order to establish a Covering Law Theorem for interrogative logic, one first has to establish an interpolation theorem for it. Here we only register the fact that such a theorem can proved, but will not try to do so.

By means of a suitable interpolation theorem we can prove a Covering Law Theorem.

THEOREM 10 (Covering Law Theorem (special case)). *Assume that the sequent*

$$\Gamma \rightarrow KP(b)$$

can be established by means of answers that do not contain P while b does not occur in Γ. Assume also that Γ does not entail $KP(b)$ and that the set of available answers is consistent. Then there is a formula $H[x]$ satisfying the following conditions:

(i) $\Gamma \rightarrow K(\forall x)(H[x] \rightarrow P(x))$
is logically valid;

(ii) $\rightarrow KH[b]$
can be established interrogatively;

(iii) Neither P not b occurs in $H[x]$;

(iv) All nonlogical constants of $H[x]$ occur both in Γ and in the available answers, except for a finite number of individual and/or function constants introduced in the course of the interrogative argument by answers to the inquirer's question.

We will not present an explicit proof of this theorem here or a proof of any of its generalizations. Such a proof would a fuller discussion of the model theory of epistemic logic that we can undertake here. It is nevertheless in order to indicate the underlying reason why the epistemic version of the Covering Law Theorem holds. It was seen in Section 4 above that the Covering Law Theorem holds when we can disregard the epistemic element. Now the *knows that* operator K is in effect merely a universal quantifier ranging over alternative models. Accordingly, it does not introduce anything that could violate the Covering Law Theorem. The operative new elements here are therefore the slashed expressions $(\exists x/K)$ and $(\vee/K)$. And they can cause havoc in the interpolation theorem (which underlies the Covering Law Theorem) in epistemic interrogative logic only when they occur in the consequent S_2 of the conditional $(S_1 \supset S_2)$ to be interpolated. The reason is that the task of interpolation means finding a separation formula I such that $S_1 \vdash I, \neg S_2 \vdash \neg I$ where $\neg$ is contradictory negation. Since I is a formula of ordinary epistemic logic, it does not cause any problems. But it is not clear how the contradictory negation of an epistemic statement can be expressed in general. However, these doubts do not affect the Covering Law Theorem, for there the consequent which is at issue there is $KP(b)$. It does not contain any slashed expressions, and hence its contradictory negation is the same as its normal (strong) negation $\sim KP(b)$, which does not take us outside the epistemic logic. Hence there is no obstacle to the use of an interpolation theorem for the purpose of proving the epistemic version of the Covering Law Theorem along roughly the same lines as in the nonepistemic case.

The epistemic Covering Law Theorem can be generalized in the same way as the nonepistemic Covering Law Theorem (special case). (See Section 4 above.)

THEOREM 11 (Strategy Theorem). *The optimal choice of the question to ask is the same as the optimal choice of the rule (L.E) of (L.$\vee f$) to apply, assuming that the question is answerable.*

Of these metatheorems, the first and the third rely on model-theoretical notions like validity. They could be taken for granted in our interrogative logic only as long as the epistemic element was not brought to bear. In the presence of the epistemic operator K, the model theory of knowledge statements need to be spelled out more explicitly.

19 ON THE LOGIC OF EXPLANATORY REASONING

According to what was indicated, in a typical interrogative reasoning the ultimate conclusion, that is, the initial sentence on the right-hand side of the initial sequent is the desideratum of the principal question which the reasoner is trying to answer. What this question is, in other words, what the reasoner is trying to find out, is indicated by the slashed ingredients $(\exists x/K)$ and $(\vee/K)$ of the desideratum.

But what happens when there are no such ingredients in the ultimate conclusion? In such a case, the reasoner is not trying to discover a new truth, an answer to a question, but to see if (and how) a pre-established conclusion follows interrogatively from given initial premises. What purpose can such an exercise have? What kind of knowledge can one hope to reach by constructing a bridge from the initial premises to the ultimate (but already fixed) conclusion?

On the interpretational level, an answer to the first of these two questions is reasonably clear. Even when we know that a statement is true, we may still want, for a good reason to know *why* it is true, in other words, we may want to explain its being true. Hence an interrogative proof of S from the initial premises T may in many cases be thought of as providing as explanation of why S is true. Why-questions and the kind of reasoning needed to answer them thus occupy a natural niche on the map of different types of interrogative reasoning. (Cf. here [Hintikka and Halonen, 1995]).

The Covering Law Theorem applies to the very case in which the ultimate conclusion does not contain any slashed expressions $(\exists x/K)$ or $(\vee/K)$. The conclusion of the Covering Law Theorem bears in fact some resemblance to what is known as the covering law model of explanation. (See [Hempel and Oppenheim, 1948; Salmon, 1990; Pitt, 1988].) T here are some marked differences, however. The suggestion we just made was to think of the interrogative proof of the explanandum from the initial premises as providing an explanation. In the strictly construed Hempelian model, transposed to our framework, an explanation is provided by a logical (not interrogative) derivation of the explanandum from the "initial conditions" by the sole means of the covering law. (In our treatment, those initial conditions cor-

respond to nature's answers to the inquirer's questions.) In our treatment of explanation, the covering law does not play any direct role. Rather, it looks merely like a mere by-product of the explanatory argument. In contrast, Hempel assigns in this approach to the covering law the same role as is played in our treatment by the initial premise T (which in scientific context means the background theory). In our view, this is a visions confusion between the respective roles of the background theory (initial premises) T and of the covering law in explanation. It is not surprising to find in many discussions of the covering law model an unmistakable assimilation of the covering law to the background theory.

Moreover, the covering law itself is available only on certain further conditions, as the Covering Law Theorem shows.

Most importantly, Hempel's covering law model does not provide any explanation of what kind of information it is that the covering law is supposed to yield. This question is obviously a variant of the more general question as to what information explanations of known facts can yield.

An answer can be obtained by having a closer look at the Covering Law Theorem. It suffices to consider the proof of this theorem in the original nonepistemic interrogative logic (see above). The antecedent $H[x]$ of the "covering law"

(53) $T \vdash (\forall x)(H[x] \supset P(x))$

is there obtained as the Craigean interpolation formula H between A and $(\wedge T^* \supset P(b))$. It has been shown by [Hintikka and Halonen, forthcoming] that (and how) H can be considered as a summary of the argument which leads from A to $(\wedge T^* \supset P(b))$. With certain unimportant exceptions, there obtains a one-to-one correlation between the logical constants in H and the applications of the different rules of interrogative reasoning in the derivation of $(\wedge T^* \supset \ P(b))$ from A. Hence the interpolation formula H codifies certain information that can be highly nontrivial. This information can be thought of as a blow-by-blow account of how the fact that the "initial conditions" expressed by A hold leads, assuming that the background theory T holds, to the result that $P(b)$. It is the character of H as a summary of the actual connection between A and the explanandum that makes H an integral part of an explanation of $P(b)$. As we may perhaps put it, the argument from A to $(\wedge T^* \supset P(b))$ explains why $P(b)$ holds by showing how the general laws codified in T lead step by step from the initial conditions A to the explanandum $P(b)$, and H provides us with a summary of all these steps in a single sentence.

When the premises of the Covering Law Theorem do not hold, the derivation of $(\wedge T^* \supset P(b))$ from A can still be considered as an explanation of sorts. In this case, it is nevertheless more natural to speak of an explanation *how* rather than of an explanation *why*. Such an account *how* cannot be summarized in any one-line explanation why.

20 EXAMPLES AND COMMENTS

In order to illustrate these remarks, we may consider a sample interrogative argument. In this example,

$$T = \text{ initial premise } = (\forall x)(\forall y)(\forall z)((Q(x)\&R(x,y)\&R(y,z)) \supset P(z)).$$

Its negation $\sim T$ is $(\exists x)(\exists y)(\exists z)(Q(x)\&R(x,y)\&R(y,z)\& \sim P(z))$.

$A =$ the set of requisite answers to questions $= \{Q(a), R(a,b), R(b,c)\}$.
$U =$ ultimate conclusion $= P(c)$.

Here the antecedent $H[x]$ of the covering law

$$(\forall x)(H[x] \supset P(x))$$

is obtained (as indicated in the proof of the covering law theorem above) as the interpolation formula from the proof of the sequent

$$A \to (T \supset U)$$

i.e.

(54) $Q(a), R(a,b), R(b,c) \to (\forall x)(\forall y)(\forall z)((Q(x)\&R(x,y)\& R(y,z)) \supset P(z)) \supset P(c))$.

A (cut-free) proof might run as follows:

(55) $A \to\sim T, P(c)$

from (54) by $(R.\vee)$.

(56) $A \to\sim T, P(c), (\exists y)(\exists z)(Q(a)\&R(a,y)\&R(y,z)\& \sim P(z))$

from (55) by (R.E).

(57) $A \to\sim T, P(c), (\exists y)(\exists z)(Q(a)\&R(a,y)\&R(yz)\& \sim P(z)), (\exists z)(Q(a)\& R(a,b)\&R(b,z)\& \sim P(z))$

from (56) by (R.E).

(58) $A \to\sim T, P(c), (\exists y)(\exists z)(Q(a)\&R(a,y)\&R(y,z)\& \sim P(z)), (\exists z)(Q(a)\&R(a,b)\&R(b,z)\& \sim P(z)), Q(a)\&R(a,b)\&R(b,c)\& \sim P(c)$

from (57) by (R.E).

At this point the argument splits into four branches (*subtableaux*). In the first three (sequents (6)–(8)) the new formula introduced to the right side is $Q(a), R(a,b)$ and $R(b,c)$, respectively. These branches close because the same formula occurs on the left side.

In the fourth case (sequent (9)), the new formula is $\sim P(c)$. The branch closes because the formula P(c) also occurs on the right side.

The interpolation formula $I_1[c]$ for (54) is obtained as indicated step by step from the argument (1)–(9) as follows:

$$
\begin{aligned}
&I_1[c] = I_2[c]\\
&I_2[c] = (\exists x)I_3[x,c]\\
&I_3[a,c] = (\exists y)I_4[a,y,c]\\
&I_4[a,b,c] = I_5[a,b,c]\\
&I_5[a,b,c] = I_6[a] \vee I_7[a,b] \vee I_8[b,c]\\
&I_6[a] = Q(a)\\
&I_7[a,b] = R(a,b)\\
&I_8[b,c] = R(b,c)\\
&I_9 = \text{ empty}
\end{aligned}
$$

Here the existential quantifiers $(\exists x)$ and $(\exists y)$ making their appearance in I_2 and I_3, respectively, are occasioned by the fact that the substituting constants a and b are imported from the left side to the right one by an application of (R.E).

Putting all the equations I_1–I_9 together we obtain as $I_1[c] = H[c] =$

$$(\exists x)(\exists y)(Q(x)\&R(x,y)\&R(y,z).$$

Hence the covering law C is

$$(\forall x)\{(\exists x)(\exists y)(Q(x)\&R(x,y)\&R(y,z)) \supset P(z)\}$$

and the "initial condition" $H[c]$ is

$$(Q(a)\&R(a,b)\&R(b,c)).$$

Here clearly the two characteristic sequents are provable, viz.

$$
\begin{aligned}
&T \to C \text{ and}\\
&A \to H[c].
\end{aligned}
$$

It can be seen where the logical constants of $H[c]$ come from. The two existential quantifiers come from the two applications of (R.E) which import the constants a and b, respectively, for the first time from the left side to the right side in steps (3) and (4). The conjunction signs come from the application of (R.$\vee$) in steps (6)–(8). The other steps in (1)–(9) do not introduce any constants into the interpolation formula.

In general, it is seen from the way in which the sequent

$$A \to (T \supset P(c))$$

is proved (so as to give rise to the interpolation formula $H[c]$) that the interpolation formula is, for all practical purposes, almost a substitution instance

of the initial premise T, unless T contains complex quantifier formulas (obtainable as answers to the reasoner's question) which can introduce new constants into the argument by means of (L.E). Unless this is the case, in other words, unless quantificationally complex propositions are available as answers to questions the covering law is little more than an instance of the initial premise. This observation throws some interesting light on the logic of explanatory reasoning. In the theory of explanation, a confusion has been rampant between the roles of the initial premise T (which in scientific explanation typically means the background theory) and the covering law $(\forall x)(H[x] \supset P(x))$ in explanation. What has been seen means that this confusion, too, can be blamed on the Atomistic Postulate. Thus the interrogative logic developed here throws some sharp light on the nature of explanatory reasoning.

21 FALLACIES

Among many other topics in the theory of applied logic our interrogative approach throws light on the theory of mistakes in reasoning, traditionally known as fallacies. (Cf. here [Hintikka, 1987].) Indeed, our theory offers an overview on different kinds of such mistakes. Corresponding to the distinction between logical inference steps and interrogative (questioning) steps, there is a distinction between such fallacies as affect logical inference steps in interrogative inquiry and such as affect questioning steps. Mistakes due to ambiguity fall into the former category, as do such outright misinferences as the mistake known as the fallacy of affirming the consequent. Mistakes in questioning include the so-called fallacy of many questions. In our framework, this mistake amount to asking a question whose presupposition has not been established. The traditional name of this fallacy is itself fallacious, however. It reflects the idea that the presupposition of the illicit question ought to have been established by a previous question, so that the mistake can be thought of as an misguided attempt to combine two or more questions ("many questions") into one. However, from our rules of interrogative logic (see Sections 2 and 15 above) it is seen that the presupposition can equally well be reached by means of a logical inference or even by a series of steps involving both inferences and questions. Hence the intended mistake cannot be thought of simply as an attempt to compress several questions into one.

An especially instructive "fallacy" is the so-called fallacy of begging the question or *petitio principii*. It is in our days usually taken to amount to assuming the conclusion of a logical inference in its premises. As has been pointed out repeatedly, such "assumption" is merely a mark of a valid deductive inference. So where is the fallacy? Here history offers an interesting sidelight on our theory. As Richard Robinson [1971] has shown, originally

the intended mistake was to ask ("petition") in an interrogative game the overall principal question instead of trying to answer it by means of a series of "small" or operative questions. The naturalness of this sense of "begging the (wrong) question" is brought out by interrogative logic. A mistake which has usually been taken to be a fallacy in inference turns out to have originally been a fallacy in questioning.

Another distinction between that is made obvious by our framework separates violations of the definitory rules of an interrogative game from violations of its strategic principles. A former kind of violation means that you are not engaged in an honest interrogative enterprise, as little as you are when pushing a chessman on the board in violation of the rules of chess is a move in chess. In contrast, a violation of the strategic rules of a game may make you guilty of stupidity but not of an infraction.

Still other kinds of fallacies involve mistaking of one kind of interrogative game for another. For instance, in one kind of game a premium is put only on the truth and informativeness of the ultimate conclusion, while in another one the sole aim is to refute another persons' thesis. Many such distinctions can in our approach be implemented by different kinds of assignments of utilities to the different courses that an interrogative game can take, including their ultimate results. Then it may turn out to be a strategic mistake in a truth-seeking game to rely on the answers of any one oracle, even in cases where the oracle is your opponent. This mistake can be considered a form of the so-called fallacy of arguing *ad hominem*.

Yet another mistake involves a confusion between the different *dramatis personae* in an applied interrogative game. For instance, it is sometimes alleged that the *ad hominem* fallacy consists in rejecting the claims a person advances simply on the basis of derogatory facts about the person making the claim. But if the person in question acts as an "oracle" in our technical sense, then it is of course according to Hoyle and not fallacious in the least to use derogatory facts about him to cast doubts on the testimony of such on oracle. In contrast, it is a mistake to try to evaluate someone's argument by reference to what one knows about the arguer.

We have a mistake here only if the role of an oracle is confused with the role of an inquirer. The mistake consists in confusing the evaluation of an oracle (e.g. a witness) with the evaluation of an arguer. An argument, unlike the testimony of witness, has to be evaluated *ad argumentum*, not *ad hominem*, to use the traditional locutions.

ACKNOWLEDGEMENT

This article represents independent work by the junior authors based on suggestions by the senior author. Ilpo Halonen has been exclusively responsible for Section 19 and Arto Mutanen for Section 9.

BIBLIOGRAPHY

[Belnap and Steel, 1976] N. D. Belnap and T. B. Steel. *The Logic of Questions and Answers*, Yale U. P., New Haven, 1976.

[Beth, 1953] E. W. Beth. On Padua's method in the theory of definition. *Indagationes Mathematicae*, **15**, 330–339, 1953.

[Beth, 1955] E. W. Beth. Semantic entailment and formal derivability. *Mededelingen van de Koninklijke Nederlandske Akademie van Wetenschappen, Afd. Letterkunde*, N. R. vol. 18, no. 13, pp. 309–342, 1955.

[Bromberger, 1992] S. Bromberger. *On What We Know When We Don't Know: Explanation, Theory, Linguistics, and How Questions Shape Them*, University of Chicago Press, 1992.

[Collingwood, 1933] R. G. Collingwood. *An Essay on Philosophical Method*, Clarendon Press, Oxford, 1933.

[Craig, 1957a] W. Craig. Linear reasoning: a new form of the Herbrand–Gentzen theorem. *Journal of Symbolic Logic*, **22**, 250–268, 1957.

[Craig, 1957b] W. Craig. Three uses of the Herbrand–Gentzen theorem in relating model theory and proof theory. *Journal of Symbolic Logic*, **22**, 269–285, 1957.

[Harrah, 1984] D. Harrah. The logic of questions. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic*, Vol. 2, D. Reidel, Dordrecht, pp. 715–764, 1984. Give details of 2nd edition as well.

[Hempel and Oppenheim, 1948] C. G. Hempel and P. Oppenheim. Studies in the logic of explanation. *Philosophy of Science*, **15**, 135–175, 1948.

[Hintikka, 1976] J. Hintikka. The semantics of questions and the questions of semantics. *Acta Philosophica Fennica*, **28**, no. 4), Societas Philosophica Fennica, Helsinki, 1976.

[Hintikka, 1984] J. Hintikka. The logic of science as model-oriented logic. In P. Asquith and P. Kitcher (eds.), *PSA 1984*, 1, Philosophy of Science Association, East Lansing, pp. 177–185, 1984.

[Hintikka, 1985] J. Hintikka. A spectrum of logics of questioning, *Philosophica*, **35**, 135–150, 1985.

[Hintikka, 1987] J. Hintikka. The fallacy of fallacies. *Argumentation*, **1**, 221–238, 1987.

[Hintikka, 1988] J. Hintikka. What is the logic of experimental inquiry? *Synthese*, **74**, 173–188, 1988.

[Hintikka, 1989] J. Hintikka. The role of logic in argumentation. *The Monist*, **72**, 3–24, 1989.

[Hintikka, 1991] J. Hintikka. Towards a general theory of identifiability. In James H. Fetzer *et al.*, editors, *Definitions and Definability*, Kluwer Academic Publishers, Dordrecht, pp. 161–183, 1991.

[Hintikka, 1992] J. Hintikka. The interrogative model of inquiry as a general theory of argumentation, *Communication and Cognition*, **25**, 221–242, 1992.

[Hintikka, 1995] J. Hintikka. The games of logic and the games of inquiry, *Dialectica*, **49**, 229–249, 1995.

[Hintikka, 1996a] J. Hintikka. On the development of Aristotle's ideas of scientific method and the structure of science. In William Wians, editor, *Aristotle's Philosophical Development: Problems and Prospects*, Rowman and Littlefield, Lanham, Maryland, pp. 83–104, 1996.

[Hintikka, 1996b] J. Hintikka. *The Principles of Mathematics Revisited*, Cambridge U. P., 1996.

[Hintikka and Halonen, 1995] J. Hintikka and I. Halonen. Semantics and pragmatics for why-questions. *Journal of Philosophy*, **92**, 636–657, 1995.

[Hintikka and Halonen, forthcoming] J. Hintikka and I. Halonen. Interpolation as explanation. Forthcoming.

[Hintikka and Harris, 1988] J. Hintikka and S. Harris. On the logic of interrogative inquiry. In A. Fine and J. Lepkin (eds.), *PSA 1988*, **2**, Philosophy of Science Association, East Lansing, pp. 233–240, 1988.

[Hsiao, 1983] C. Hsiao. Identification. In Z. Griliches and M. D. Intriligator, editors, *Handbook of Econometrics*, vol. 1, chapter 4, North-Holland, Amsterdam, 1983.

[Piatelli-Palmarini, 1994] M. Piatelli-Palmarini. *Inevitable Illusions*, John Wiley and Sons, NY, 1994.

[Pitt, 1988] J. C. Pitt, editor. *Theories of Explanation*, Oxford University Press, New York, 1988.

[Robinson, 1963] A. Robinson. *Introduction to Model Theory and the Metamathematics of Algebra*, North-Holland, Amsterdam, 1963.

[Robinson, 1971] R. Robinson. Begging the question 1971. *Analysis*, **31**, 113–117, 1971.

[Salmon, 1990] W. C. Salmon. *Four Decades of Scientific Explanation*, University of Minnesota Press, Minneapolis, 1990.

[Wisniewski, 1995] A. Wisniewski. *The Posing of Questions: Logical Foundations of Erotetic Inferences*, Kluwer Academic, Dordrecht, 1995.

R. H. JOHNSON AND J. A. BLAIR

INFORMAL LOGIC AND THE RECONFIGURATION OF LOGIC

I. Introduction

1 AIMS OF THE CHAPTER

This chapter describes how informal logic emerged as an attempt to reconfigure one part of logic in the wake of the sorts of critiques discussed in Chapters 3 and 4. Informal logic appeared first as an attempt to reform the introductory-level undergraduate logic course [Kahane, 1971; Scriven, 1976]. That project quickly took the form of attempts to capture what might be called the logic of everyday argumentation [Govier, 1987], with as well some reference to the logic of the stylized discourse of the disciplines [Weinstein, 1990]. In its most recent incarnation, informal logic has taken the form of developing wide-ranging theories of argument, with particular attention to the normative issues that any such theory must address.

2 CONTENTS OF THE CHAPTER

- the origins of informal logic: how it first emerged and then developed (Section II);
- the nature of informal logic: its historical antecedents, misconceptions, our own conception, and other prominent perspectives (Section III);
- the important theoretical issues and the various approaches to them: informal logic in the attempt to develop criteria, methods and procedures for the construction and evaluation of arguments, whether in ordinary (or mundane) discourse, or in the more stylized discourses of the disciplines — including a discussion of where the study of fallacies belongs in the informal logic project (Section IV);
- the presentation of a research agenda for future work informal logic (Section V);
- the resources currently available to researchers and the discipline (Section VI);
- the implications of informal logic (Section VII);
- a summary of the chapter (Section VIII)

II. The Origins of Informal Logic

3 INTRODUCTION

Informal logic may be seen as another chapter in the ongoing story of logic reforming itself [Johnson and Blair, 1985]. Informal logic developed in response to a set of critiques of what had become the dominant form of logic in the 20th century — formal deductive logic. What were the critiques, as its originators saw them, which led them to call for such reform?

There were three related sorts of critique. The first arose from the frustrations encountered in trying to improve undergraduates' logical thinking by teaching them formal deductive logic. These frustrations crystallized around what might be called the "pedagogical critique", to be discussed below. The second developed out of problems within or challenges to logic itself, or at least some philosophers' conception of logic, and it might be called the "internal critique." Its problems have been discussed in Chapter 3. The third emerged from empirical observations of reasoning and argumentation and might be called the "empirical critique." It was discussed in Chapter 4. We now comment on each of these in turn.

4 THE PEDAGOGICAL CRITIQUE

To understand the development of informal logic, it is necessary to understand the context in which it arose. That context may be called pedagogical in the first instance, rather than theoretical, if 'pedagogy' is taken to refer to the means used to realize an instructional goal.

The pedagogical critique arose from the emerging belief that traditional formal deductive logic does not provide the student with adequate tools for the analysis and the evaluation of argumentation as that occurs in ordinary discourse. (Later it will be argued that it does not do the job for discourse within the disciplines.)

Who makes this critique? Many did, but one of the first among them was Howard Kahane, who was himself the author of a standard formal logic textbook [Kahane, 1969] and so someone who cannot be considered unsympathetic to traditional logic. Here is what Kahane wrote in 1971:

> Today's students demand a marriage of theory and practice. That is why so many of them judge introductory courses on logic, fallacy and even rhetoric not relevant to their interests.
>
> In class a few years back, while I was going over the (to me) fascinating intricacies of the predicate logic quantifier rules, a student asked in disgust how anything he'd learned all semester long had any bearing whatever on President Johnson's decision

> to escalate again in Vietnam. I mumbled something about bad logic on Johnson's part, and then stated that Introduction to Logic was not that kind of course. His reply was to ask what courses did take up such matters, and I had to admit that so far as I knew none did.
>
> He wanted what most students today want, a course relevant to everyday reasoning, a course relevant to the arguments they hear and read about race, pollution, poverty, sex, atomic warfare, the population explosion, and all the other problems faced by the human race in the second half of the twentieth century. (vii)

The term 'relevant' may be problematic here. What Kahane seems to have had in mind, given the contrast between his formal logic textbook [1969] to which the student was responding, and the new kind of textbook in the Preface of which these remarks appeared, is that learning such things as the predicate logic quantifier rules in abstraction from any application does not help in evaluating political and social issues of the day, and that the kind of instructional material that is needed for such application is different from predicate logic. There also seems to have been a background assumption that the reason many students enrol in "logic" courses is to improve their practical logical skills, and that studying the rules of FDL does not appear to accomplish that goal. In sympathising with this kind of student demand, then, Kahane wants logic to be useful for the purposes of social and political critique, to be relevant to the mundane argumentation surrounding the public controversies of the day. Learning formal deductive logic is judged not to be useful for those purposes. A number of reasons were identified.

In the *first* place, arguments "on the hoof" [Woods, 1995] are expressed with what Johnson [1981] calls "clutter". The argument as a set of reasons supporting a claim will be formulated using various rhetorical and stylistic devices; it will be accompanied by explanations of its terms and by background information relating it to its context; there will be repetitions for emphasis and other extraneous material (asides arising from associations by the arguer that play no argumentative role). In order to employ standard propositional or predicate logic, all of these "overlays" will have to be pared away before the propositional core of the argument is reached and made ready for re-expression in canonical notation. But that process of removing the argumentative flesh from its formal structural bones itself requires the exercise of critical logical judgement and it also risks the removal of substantive material that is germane to the argumentation.

Second, once the "extraneous" material has been purged, there remains the often significant and non-trivial step of encoding the argument into the canonical notion. This step is not itself formalized, and is arguably not formalizable [Bar-Hillel, 1969; Woods and Walton, 1972], so formal deductive

logic applied to arguments on the hoof arguably must depend upon some other theory to guide this task of translation.

Third, suppose that the argument has been translated into the notation of, for example, propositional logic. In most cases, the argument will be invalid as it stands. The deployment of formal deductive logical tools thus faces a dilemma. Grasping one horn implies judging that most arguments on the hoof are invalid, full stop. Yet somehow the world carries on, and not entirely irrationally. So the application of logic does not exhaust, or fully illuminate, the operations of rationality. Grasping the other horn implies judging that most arguments on the hoof are somehow incomplete, and that they are indeed valid once their missing parts are supplied, or their assumed or unexpressed components are made fully explicit. But having thus helped in rendering the argument valid, the tools of formal deductive logic have completed their task, leaving behind the interesting substantive question of whether the material supplied to render the argument valid is true. Thus, for such deductive reconstructions, the argument's problems will always turn out to be substantive questions about the truth or reasonableness of the premises, either explicit or "missing", and never logical questions about its formal structure. That leads to a fourth point.

Fourth, traditional logic recognized only deductive argumentation. If one believes that there are logically good arguments that are not deductive in character, then formal deductive logic will have to be supplemented by some other logical theory.

The result is the following situation. The student is required to master a formidable set of symbolic techniques and machinery, which are then with great difficulty and sometimes some sleight of hand applied to actual arguments in natural language, with a net result that is usually logically unilluminating. In other words, a great deal of work is done for little profit.

Involved as well is the gap between the ordinary intuitions about the meaning of "if-then" and the story about them told by propositional logic. Thus, according to a standard interpretation of the classical propositional calculus, the following conditional must be accepted: "If Fermat's Last Theorem is true, then either Fermat's Last Theorem is true or God exists". This conditional is sanctioned by the classical propositional calculus. Not only students find it very difficult to accept this result. Of course logicians have tried in various ways to make it acceptable, and the classic dissatisfaction has led to the development of model logic and relevance logics.

Now it is possible to think that these problems are really not trenchant and that the abandonment of formal deductive logic as the tool of choice for the undergraduate critical reasoning course on these grounds is simply the result of the teacher's having experienced a failure of nerve. Or, one might think that these difficulties occur only because the students are not up to the intellectual challenge of "real" logic, and so the replacement of formal logic in the introductory reasoning course by a kind of watered down,

"baby" logic, namely informal logic, is (another!) symptom of the erosion of educational standards. We note these views at this point to acknowledge their existence. Our grounds for disagreeing with them have been stated in part already, and will emerge more fully as we proceed.

For the moment, however, it is useful to try to understand how the pedagogical situation giving rise to informal logic came about. To do so we must take into account both the history of logic and logic teaching in this century, and also the cultural context in which the informal logic critique appeared: that is, English-speaking North American undergraduate universities in the 1960s. We now turn to these matters.

Logic textbooks in the 20th century

To get a sense for the pedagogical developments leading up to informal logic, it is instructive to compare textbooks from different periods in the 20th century.

We begin with H.W.B. Joseph's *Introduction to Logic*, published in 1906. In this textbook Aristotelian syllogistic was still the dominant logical theory. The first four chapters deal with the syllogism and when Joseph turns (Chapter 15) to hypothetical and disjunctive reasoning, he treats both syllogistically. Chapters 18 through 24 cover various aspects of inductive reasoning including causal reasoning, explanation and reasoning by analogy. Chapter 25 is on mathematical reasoning and Chapter 26 on the methodology of the sciences. There is an appendix on fallacies which, following Aristotle, divides fallacies into those *in dictione* (equivocation, amphibole, composition, division, accent and figure of speech) and those *extra dictionem* (accident, *secundum quid, ignoratio elenchi, petitio principii, non causa pro causa*, the consequent, and many questions). While the term 'inference' has a listing in the index, the term 'argument' does not.

If we look next at L. Susan Stebbing's *A Modern Introduction to Logic* (1931), we notice a quite different alignment. This textbook has three parts. Part I deals mostly with deduction: both with traditional syllogistic relationships and patterns and with elements of propositional logic. Part II, taking up such topics as induction by enumeration and by analogy, causality and hypothesis formation, can be said to deal mostly with induction. Part III covers definition, abstraction and generalization, the characteristics of logical thinking and finally the history of logic. Although Stebbing states explicitly that the book is to be an introduction to logic *simpliciter*, not to symbolic logic or mathematical logic, it shows the influence of *Principia Mathematica* while maintaining the utility of traditional Aristotelian logic. In other words, taking Joseph and Stebbing as exemplars, in the textbooks deductive logic was expanded to include the recent developments in mathematical or symbolic logic: syllogistic became complemented by propositional logic.

Next we consider a post-war textbook, Max Black's *Critical Thinking* [1946]. Its three sections were: Part I, Deductive Logic (including a chapter on the syllogism); Part Two, Language (containing chapters on the uses of language, ambiguity, definition and assorted fallacies); and Part Three, Induction and Scientific Method. Notice that Black used the title "Critical Thinking" for what otherwise might be called an introductory logic text, but what is most significant is that propositional logic has by this time become the dominant theory of argument.

Turning to W.V. Quine's *Methods of Logic* [1952], we see a complete change. For Quine, not arguments or inferences but rather relations between statements are the focus of logic. He wrote "the chief importance of logic lies in implication, which, therefore, will be the main theme of this book" (xvi). The entire book is taken up with the techniques of modern symbolic logic. A reader familiar with Quinean themes will find them reflected in the Introduction: the importance of statements, the nature of logical truth, the under-determination of systems by experience, and so on. But Quine did not encourage the student or teacher to expect enlightenment from *Methods of Logic* about how to appraise the editorial in the morning newspaper, and none is forthcoming in this textbook. (In fact, Quine's contribution to training undergraduates in reasoning and argument is to be found not in a logic textbook, but rather in Quine and Ullian's applied epistemology textbook, *The Web of Belief* [1970].)

Last, but certainly not least, we come to Irving M. Copi's *Introduction to Logic* (1st edition 1954) which embodied what we have called the "global" approach [Blair and Johnson, 1980]. Like Black's Critical Thinking, it contains three parts: Part One, about language; Part Two, about deduction; and Part Three, about induction. For whatever reasons, Copi's textbook caught on and became the standard against which other logic textbooks would have to array themselves.

If this thumbnail sketch of selected textbooks is representative, and we submit that it is, there can be seen in the development of logic textbooks in the 20th century the increasing influence of formal deductive logic as the theory in charge. Inductive logic is second in command, and one may find a section or an appendix in which the author deals in passing with the fallacies. The typical introductory logic textbook and the course in which the student received introductory logic instruction (the "baby logic" course) were predicated on the assumptions that logic is deductive logic, and deductive logic is the logic of reasoning and argument. The situation did not change for more than a decade and a half (the years of our own undergraduate and graduate education in logic, by the way), until the appearance of Kahane's *Logic and Contemporary Rhetoric* in 1971, the textbook that launched the informal logic movement. Kahane's was the first of what we call the "new wave" introductory logic course textbook.

Speculations about the cultural origins of informal logic

The following comments consist mostly of empirical claims, none of which has been subjected to the scrutiny of sociological investigation. We thus can offer them as no more than hypotheses. We would welcome their systematic investigation by researchers in the history of ideas or the sociology of knowledge who might be interested in testing them.

During the 1960s, university students in the United States and Canada had taken an increasingly political stance, not only towards the university but towards the culture itself. Some students spent their summers participating in the anti-segregation civil rights voter-registration and protest marches in the American south, and brought their concerns back with them to their campuses in the autumn. This was also the period of the active anti-Viet Nam war protest movement. In the universities, students were demanding that their courses relate to their felt needs as citizens critical of the status quo. This activism, kindled by the civil rights movement early in the decade and in turn fanned by the anti-Vietnam war movement and the related demand for "relevance" in education associated with the mid- and late-1960s, had its impact on student expectations when they entered their classrooms. They expected their introductory logic course to help them to interpret and assess the reasoning and argumentation about current affairs. The texts of the 1950s and 1960s, following Copi's (1st edition, 1953), did not satisfy their needs. At the same time, many instructors who also wanted logic to be useful in these ways were sympathetic. The existence of such a cultural *Geist* seems to be the best explanation for what we witnessed in the early 1970s the emergence, almost at the same time, but independently of one another, of course syllabi and soon thereafter textbooks that replaced, or supplemented, the standard introduction to symbolic logic course with one designed to teach how to assess critically the logic of natural language argumentation in public discourse. Such courses proved to be extremely popular with students.

Part of the phenomenon of the emergence of the "radical" 1960s out of the docile 1950s was a general challenge to the more authoritarian methods of education then current, and it found vocal expression in the universities and colleges. Students expressed distrust at being required to accept on faith the contents of the curriculum and the authority of their instructors, and began to challenge both. Many instructors welcomed this decline of student passivity and the increasing critical involvement by students in their own education. In *Logic: The Theory of Inquiry* [1939], Dewey had argued that logical forms, and indeed logic itself, develop in response to the changing environment. Informal logic can be seen as one attempt to adapt logic to the changing environment in higher education in North America.

Simultaneously with the emergence of informal logic, there also occurred a revival of the Deweyan ideal of reflective thinking under the rubric "critical

thinking". In *How We Think* [1910], a work explicitly devoted to training in thinking, Dewey formulated one of the first, and an influential, characterization of such thinking:

> *Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it, and further conclusions to which it tends, constitutes reflective thought. (1991, 6: italics in the original)*

(See also Dewey's *The Quest for Certainty*, 1929, Ch. V.) This idea was never far from the educational mainstream in any case, as Max Black's post-war introductory logic textbook title, "*Critical Thinking*", attests. If critical thinking includes at its core, along with the critical assessment of knowledge claims, also the critical appraisal of reasoning and arguments, and if informal logic is understood as having the critical appraisal of living arguments as its subject-matter, it is easy to see how a movement for the development of critical thinking could not only reinforce, but even become identified with, a movement for the development of an informal logic. This, we submit, is what happened. It is our contention, therefore, that there are two quite distinct strands, which it is often important to disentangle. The terms 'critical thinking' and 'informal logic,' which are often treated as equivalent, denote critiques that, although consistent and related, are different.

In addition, by coincidence in the late 1970s there was a demographically-induced drop in post-secondary enrolments in the United States and Canada, with the result that programs seeking to maintain their levels of staffing-when staff allotments were related to teaching loadsfound themselves having to compete against one another for students within the university or college. In this effort to attract enrolments, philosophy departments entered the competitive fray by, among other ways, sponsoring the new "applied" or "practical" logic course that had already been found to be much more popular than the historical introductory formal logic course, and they presented it as a course in "critical thinking" or "critical reasoning". Publishers were fairly quick to recognize this new market, and so by the 1980s there was a proliferation of textbooks solicited by sales-conscious publishers, the promotion of which probably contributed to even more of these new courses.

Kahane's 1971 textbook can be seen as one type of pedagogical response to this demand, and its enormous success (8th edition 1997) is evidence of the demand. He took as his focal point what had been a marginal player in the traditional introductory logic course — the part on fallacies. In making this move, he may be represented as going back to Aristotle's *De Sophisticis Elenchis* rather than the Prior Analytics. And though Kahane did not call this approach to logic "informal logic", it is natural enough that others soon would do so because of the phrase "informal fallacies" (see Fogelin's 1978 subtitle: *An Introduction to Informal Logic*). But Kahane did much more

than return to the fallacy tradition; he endeavoured to breathe new life into it. He purged the "hoary" invented examples of the fallacy tradition and replaced them with actual examples taken from contemporary discourse on political and social issues. He dropped from his list classical fallacies that are likely to deceive no one, and added new ones. So one will not find any treatment of the classic fallacy known as amphibole. One will find that, in addition to *ad hominem*, there is *appeal to authority*; and besides hasty conclusion and *ad hominem*, there are *suppressed evidence*, *red herring* and *provincialism*. Moreover, Kahane replaced the traditional chapters on scientific method and probability with chapters on the mass media, advertising, the news media and reading textbooks critically. These topics are much closer to the interests of those who wish to deal with "contemporary rhetoric".

Kahane's text was an important breakthrough, and it helped pave the way for what came to be known as "the informal logic movement", but it is important to understand that Kahane was not alone in challenging the conventional assumptions about logic tuition as found in logic texts. Earlier, Monroe Beardsley's *Practical Logic* [1955] and *Thinking Straight* [1966], then Stephen N. Thomas's, *Practical Reasoning in Natural Language* (1st edition 1973), Michael Scriven's *Reasoning* [1976] and Robert J. Fogelin's *Understanding Arguments: An Introduction to Informal Logic* (1st edition 1978) all presented alternatives to the traditional approach to the teaching of logic reasoning and argument interpretation, analysis and critique.

5 THE INTERNAL CRITIQUE

In addition to, and to some extent growing out of, the increasingly widely-shared belief about the pedagogical limitations of formal deductive logic as a tool for teaching argument appraisal and critical thinking in general, there emerged some challenges at the theoretical level to formal deductive logic construed as the theory of argument. We call this the "internal critique". Some argued that formal deductive logic is inadequate as a normative theory of argument, supplying neither necessary nor sufficient conditions for a logically good argument. Others held that it is a category-mistake to conceive of formal deductive logic as a candidate for a theory of argument. In his chapter, Harman discusses the claim that formal logic should not be construed as a theory of reasoning either.

According to the standard application of formal deductive logic to the evaluation of arguments, an argument is a good one if and only if it is "sound", that is, it is valid and its premises are true. It follows that the conclusion of a sound argument must be true. Hence, if this account is correct, there cannot be good arguments both for and against a given proposition (see [Hamblin, 1970]), for the conclusion of such arguments would be con-

tradictory. Yet it seems incontestable that there can be good arguments for and against a given proposition, such as a recommendation for an action or a social policy, or a verdict in a criminal trial, or an historical claim. These are all contexts in which frequently a strong case, that is "good" arguments, can be made both for and against. Unless there are *a priori* grounds for rejecting such arguments as good ones, it follows that soundness cannot be a necessary condition of a good argument. Furthermore, it cannot it be a sufficient condition either, because valid arguments with true premises that beg the question (for instance, that have the form $p \supset p$) are sound, but they are not good arguments — or, at least, not good arguments for the truth or for the acceptability of the proposition that is their conclusion.

The theoretical critique of formal deductive logic *as a theory of argument* pertains to each of the two elements of formal deductive logic. There is, first, the challenge to deductivism, and there is, second, the challenge to formalism.

5.1 The challenge to deductivism

Traditional logic contains a bias in favour of deductive reasoning — the view that goes back to Plato, is furthered in the Cartesian project and survives in the claim (whose source we cannot locate) that "all inference is either deductive or defective", and which we will call "deductivism". For example, Kneale and Kneale's magisterial historical study titled, *The Development of Logic* (1962), is in fact exclusively a history of the development of *deductive* logic. The two were seen as coextensive.

Problems with the validity requirement

It is perhaps worth noting that deductivism has come under fire from other quarters besides informal logic, specifically from some philosophers of science. (See Grünbaum and Salmon, *The Limitations of Deductivism*, 1988, where Grünbaum (xv) also attributes the origin or the pejorative label, "deductive chauvinism", to J. Alberto Coffa). In fact, the first challenge to deductivism is to be found in the development of inductive logic. That is, it is recognized that inductive arguments cannot be valid, for in an inductive argument the conclusion does not follow from the premises necessarily but only with some degree of probability. Yet inductive arguments can be logically "good" even though not sound. Whether or not there is a logic of induction is a matter of controversy. Some, such as Popper [1959], argue that induction understood in any classical way is not justifiable; others, such as Harman [1986], argue that while inductive reasoning may be justified in some sense, there is no "logic" of induction. Furthermore, among those who do take it that there is a logic of induction, there are disagreements over whether the fundamental notion of prob-

ability is to be construed objectively or subjectively (see [Cohen, 1995; Eells, 1995]). And there is as well Cohen's challenge to what might be called the standard logic of induction [Cohen, 1989].

The challenge presented by inductive logic has led many to take the modified view, which Govier [1987] has called "positivism", that every good argument must be either deductively valid or inductively sound [Skyrms, 1966]. Govier discusses problems with this view, the principal one being that there seem to be perfectly good arguments that are *neither* deductively sound *nor* inductively strong (in any standard sense of empirical induction). Following Wellman [1971], Govier refers to this sort of argument as "conductive", and she discusses two species of it: arguments from *a priori* analogy (as distinct from arguments from inductive analogy) and case-by-case reasoning. The former are typically normative arguments in which it is reasoned that because an action of a given kind was taken in circumstances of a give sort, so in similar circumstances another action of that kind is justified (for example, since a professor gave one student an essay-extension because of illness, so too should the professor give a similarly-situated student an essay-extension). Typical of case-by-case arguments are arguments about hiring or product-evaluation arguments. Here we have types of argument which do not seem to be reducible to either the deductive or the inductive type, yet which clearly can be logically good.

Toulmin's [1958] critique of logic was another challenge to the traditional deductivist view. Toulmin complained about the long-lasting influence of the geometrical model in which the conclusion is supposed to follow *necessarily* from the premises, as a conclusion in a proof. As a consequence of this critique he developed an original and widely-adopted approach to modelling the structure of arguments. We discuss Toulmin's model further below but for now the noteworthy point is that Toulmin is another theorist opposed to deductivism and the validity-requirement.

At the same time, but independently of Toulmin, Perelman and Olbrechts-Tyteca [1958] in their own way argued for a limited role for deductivism. They contended that formal deductive logic applies to demonstration, which they conceived as reasoning by necessary steps from self-evident truths to their logical implications. Such matters are not subject to dispute. But, they held, if one is dealing with argumentation, a domain in which claims are contestable, then the reasoning is not deductive — and yet it is, or can be, reasonable.

The work of both Toulmin and Perelman and Olbrechts-Tyteca influenced informal logic later in the day. While a few informal logicians read their work in the 1970s, the scholarly community that took it most seriously and was considerably influenced by it in the 1960s and early 1970s (Perelman and Olbrechts-Tyteca's book was translated into English only in 1969) was speech communication in the United States. After speech communication and informal logic scholars began to make contact and read one another's

work in the 1980s, the latter were far more apt to turn to Toulmin's and Perelman and Olbrechts-Tyteca's 1958 monographs.

Problems with the truth-requirement

There are problems as well with the truth-requirement. Hamblin [1970, Chapter 7] argued that truth is neither a necessary nor a sufficient condition for a good argument. He held that truth is not sufficient because it is not the truth of the premises, as from a God's-eye-view, that is pertinent, but rather whether the arguer and addressee know (or have good reason to believe) them to be true. Nor did Hamblin think truth is necessary in every case of good argument, because at least for certain practical purposes, the relevant issue is whether the arguer and the addressee accept the premises, even if they do not know them to be true. Hamblin's critique has led to one of the main problems occupying informal logic in recent years: *the formulation of a theory of premise-adequacy.*

5.2 The challenge to formalism

Another critique of deductivism is the challenge to formalism, since the deductivism and formalism are linked if an argument is said to be deductively valid if and only if it exemplifies a valid form of argument. (See [Massey, 1981], for challenges to this assumption. Also, see [Massey, 1998], for a criticism of the challenges to formalism.)

What were the concerns of those who objected to formalism? We have in mind the three problems detailed above in discussing the pedagogical critique. The first is, in many cases, the problematic character of the translation of arguments originally expressed in natural language into canonical notation. The "clutter" has to be pared away and ambiguity and vagueness removed, but that often requires sacrificing material that seemed intended to be part of the argument. Thus it can become questionable how close the resemblance is between the argument as originally expressed and its re-expression in the formal notation. Second, there is the whole problem with logical form (see [Massey, 1981]). Second, there are problems with using logical form to determine validity/invalidity. A given argument may instantiate both a valid and an invalid logical form (see [Massey, 1981]). Third, once the argument is reconstructed, it is often invalid. Although it is simple to render the argument valid by inserting the "associated conditional" a conditional with all the other premises as its antecedent and the conclusion as its consequent as an unexpressed premise, the result is that an argument's reasoning can never be suspect, and all substantive assessment of the argument must focus on the merits of its premises. In short the payoff from invoking formal criteria for argument evaluation seems disproportionately small for the amount of work required.

These, then, were the concerns that caused many to consider formal deductive logic inadequate as a theory of argument.

5.3 The empirical critique

Beginning with the work of Wason and Johnson-Laird [1972] in the 1970s, another type of critique began to emerge. Since this critique has already been discussed in large measure by Perkins, in Chapter 4, and because it was not a significant influence on the development of informal logic, we will not discuss it here. The empirical critique took place against the background of the belief (or assumption) that formal deductive logic is the correct theory of reasoning, and that learning formal deductive logic will improve one's reasoning ability. The former is a theoretical claim, and we will discuss it later. The latter is an empirical claim, and there are empirical (psychological) findings that suggest that even those who have had training in formal deductive logic do not, in general outperform on inferential reasoning tasks those who have not had such training. Indeed, those who have formal deductive logic training tend to fail at a test of basic conditional reasoning called "the selection task" (see [Finocchiaro, 1980], for a discussion of some of these issues). Other research on the effectiveness of logic instruction is mixed (see [Lehman *et al.*, 1988; Nisbett, 1993]).

6 SUMMATION

The informal logic movement began as a change in the way the introductory undergraduate logic course was taught in American and Canadian universities. It first appeared as a set of important shifts in the textbooks of the day. The changes were primarily driven by pedagogical concerns, and secondarily by theoretical problems that the methodology of formal deductive logic presented. As well, it appears that cultural conditions at the time created a receptivity for the pedagogical and theoretical critiques.

At the time these shifts occurred, there was no journal where the emerging theoretical issues could be thrashed out. Later it became clear that the informal logic movement had other and higher aims than simply changing how students were taught to interpret and assess arguments, as though one could simply change instruments without at the same time making significant adjustments elsewhere. It gradually became apparent that informal logic was concerned with reforming logic itself [Johnson and Blair, 1985] by adding a new domain which was more attentive to the needs of argumentation analysis and critique, and reminding logicians of the art and craft of logic — views which have their roots in Aristotle.

In this way the development of informal logic presents a contrast with

that of formal logic, which began as a revolution at the level of theory that later filtered down into logic textbooks. Admittedly, the latter order is the normal one: textbooks are instruments for passing the latest theory on to the next generation and hence the theoretical developments found in textbooks normally represent the precipitate from theoretical controversies that occurred earlier in the journal and monograph literature. This is the orientation Woods reflects when he wrote: [1992, 44]: "Get the theory right first, then take care (if you know how) of instructional routines". It is also the view that Massey assumes when he claims that textbooks should be driven by theory [1981].

However, informal logic has displayed a different course of development, one that might be summarized along the following lines: "Be clear about what tools are needed — or need to be designed — to do the job properly, then develop the theory needed to support the toolkit". Something like this pattern seems to have occurred in the emergence of informal logic, for developments at the theoretical level were initially largely motivated by the attempt to teach students how to assess arguments in use: how to understand and display their structure; and how to judge their relative merits.

Having discussed the emergence of informal logic, we turn next to a more systematic account of its nature.

III. The Nature of Informal Logic

1 INTRODUCTION

In their recent anthology about the history of informal logic, Walton and Brinton [1997] say that, "Informal logic has yet to come together as a clearly defined discipline, one organized around some well-defined and agreed upon systematic techniques that have a definite structure and can be decisively applied by users" (9). We agree, so in giving our account of the origins of informal logic, we will perforce be employing our own conception and differentiating it from others. In the process we will also distinguish informal logic from formal deductive logic, epistemology and critical thinking.

2 HISTORICAL ANTECEDENTS

In logic, as in so many other areas of philosophy, all roads lead back to Aristotle. While *Prior Analytics* is typically identified as the first work of formal logic, it is customary to trace informal logic back to the *Topics*, where Aristotle studies dialectical reasoning; to the *De Sophisticis Elenchis*, where he originates the fallacy tradition; and to the *Rhetoric*, where

he studies the social role of argumentation. In these three works are to be found the elements — applied or practical argument assessment and critique, argumentation as dialectical interchange, and argument as situated in rich social contexts — which in a general way characterize the orientation of informal logic as we understand it. Thus Aristotle can be said to have laid down the basis for informal logic no less than for formal logic.

In the recent history of logic, formal deductive logic has received the closest attention from logicians. Its impressive theoretical achievements have been widely and justly celebrated. By comparison informal logic has been at best a character waiting in the wings. Although it is not accurate to say that nothing of importance happens in informal logic between Aristotle and the 20th century (see [Walton and Brinton, 1997]), nevertheless the emergence of informal logic is a late 20th century development.

To our knowledge, the term 'informal logic' first appears in something like the sense in which we are using it here in 1964, in Nicholas Rescher's *Introduction to Logic*, and Carney and Scheer's *Introduction to Logic*. Neither party gives any definition of the term. We consider each in turn briefly.

Rescher. Part I of Rescher's text is titled "Informal Logic". (Part II focuses on syllogistic logic, Part III on symbolic logic — propositional and quantificational — and Part IV on inductive logic.) In the introduction, Rescher characterizes Part I as being about informal language and common discourse logic, and it covers these topics: Ch. 1, on logic and discourse; Ch. 2, on words, names and terms; Ch. 3, on definition and classification; Ch. 4, on evidence, argument and fallacies; Ch. 5, on informal fallacies; and Ch. 6, on logical exposition. 'Informal logic' thus appears to be a term used by Rescher to refer to a collection of topics, foremost among which are language and its role in logic, plus the informal fallacies.

Carney and Scheer. Here 'Informal Logic' is used without any explanation, as the title of Part One of the textbook, containing the following six chapters: Ch. 1, on appraising arguments logically; Ch. 2, covering traditional informal fallacies (divided into fallacies of relevance, insufficient evidence, and ambiguity); Ch. 3, devoted to definitions; Ch. 4, on uses of language; Ch. 5, about analogy; and Ch. 6, about dilemmas and paradoxes. Part 2 of the book is called "Formal Logic" and Part 3 is called "The Logical Structure of Science". Although Carney and Scheer do not say what exactly they understand by the term 'informal logic,' here is how they describe Part I:

> Part I, Informal Logic, contains both new and traditional topics and some novel treatment of traditional topics. Informal fallacies, which are useful in motivating students in the study of logic, are treated in the traditional manner. The standard topics — analogy, dilemmas, uses of language, classifications of methods of defining terms and types of definition — are also discussed in this part. But two topics are introduced — paradoxes

> and nonsense — which are not found in many introductory logic textbooks. The discussions of nonsense and the uses of language are designed to bring to the attention of the student some of the immediate connections between logic and language and to show the student the significance of logic to philosophy. (vii)

While Carney and Scheer believe that there are traditional topics for informal logic, it is unclear what tradition they have in mind. Their innovations are aimed at showing the student the significance of logic to philosophy, not at the appraisal of reasoning in everyday situations. It looks to us, on the basis of their description and the contents of Part One, that what Carney and Scheer mean by 'informal logic' is, besides the informal fallacies, "these matters related to logic not taken up by formal logic". This may be the first expression of the idea that informal logic attends to those matters left over from, or not attended to, or propaedeutic to, formal logic: such matters as encoding the argument in the canonic notion of formal deductive logic, or the role of language, for example.

Thus Rescher's use of 'informal logic,' differs from that of Carney and Scheer. He seems to take the term to refer to the informal fallacies and matters of language, whereas they seem to think it denotes matters not taken up by formal logic, including how formal logic applies to actual examples. Such differences in the use of 'informal logic' continue to be found as the term gains currency.

The next use we are aware of is that of Fogelin, who employs 'informal logic' as the subtitle of his textbook, *Understanding Arguments* [1978], where he writes (v–vi):

> For certain purposes arguments are best studied as abstract patterns. . . . The task of logic is to discover the fundamental principles for distinguishing good arguments from bad ones. The study of those general principles that make certain patterns of argument reasonable (or valid) and other patterns of argument unreasonable (invalid) is called formal logic.
>
> A different but complementary way of viewing an argument is to treat it as a particular use of language: arguing is one of the things that we do with words. This approach places stress upon arguing as a linguistic activity. . . . It raises questions of the following kind: What is the place of argument within language as a whole? In a given language, what words or phrases are characteristic of argument? What task or tasks are arguments supposed to perform? When an approach to arguments has this form, the study is called informal logic.

Fogelin goes on to mention the work of Grice and Austin as significant. It appears, then, that by 'informal logic' Fogelin means the study of the

linguistic and pragmatic aspects of argument — a view that comes close to Rescher's, and one that Walton will later defend [1989b].

We (Johnson and Blair) first used the term 'informal logic' in the title of the 1978 conference, "Symposium on Informal Logic", which we organized at the University of Windsor. Until that time, we had been using the term 'applied logic' to designate our enterprise. We believe that the term 'informal logic' recommended itself to us to designate this new approach to the teaching of logic in virtue of the then-recent renewed interest in the informal fallacies. Kahane had made informal fallacies the focal point of his revolutionary textbook, which we had used in our classes, and Woods and Walton had already begun their ambitious research project focussing on the informal fallacies. We had found ourselves engaged in teaching students how to evaluate arguments — a task associated with logic. Following Kahane, we took as our focal point the actual arguments about the issues of the day taken from current newspapers and magazines. The logic we were teaching was not a formal logic: we made no reference to the logical form of arguments in understanding their structure, and we did not make use of the standard of validity as part of our theory of appraisal, but used fallacy theory instead. Since the fallacies in question were the informal fallacies, the term 'informal logic' seemed appropriate.

In our first attempt to articulate the denotation of 'informal logic,' we noted wide range of meaning associated with the term: [Blair and Johnson, 1980]:

> The label "informal logic" means different things to different people. To many it refers to the lists of informal fallacies and various descriptions and classifications of the these fallacies. . . . To others it designates the subject matter of a certain sort of introductory logic course which employs various non-formal techniques to try to teach elementary reasoning skills. To still others it has come to mark off a field of logical investigation distinct from formal deductive logic. (Introduction, ix.)

In our paper, "The Recent Development of Informal Logic" (1980), we avoided any direct statement of what informal logic was but noted two tendencies that seemed to us to be identified with this development. The first was the turn in the direction of actual arguments as contrasted with the artificial types of argument often found in formal logic textbooks. The other was the growing disenchantment with the capacity of formal logic to provide the standards of good reasoning. We subdivided informal logic into (a) the theory of fallacies and (b) the theory of argument [Johnson and Blair, 1980, pp. 6–10].

Given these early uses, what can be said about the emergence of this term? Kahane's textbook (1971) showed one way logic could be revitalized,

both as a subject matter to be studied and a subject matter to be taught: by dipping back into the origins of logic, focus on the informal fallacies, and use them as in effect a theory of criticism for logic. Hence the term "informal logic". Later, the label came to designate any approach to logic which either avoided or minimized the use of formal logic as the theory of analysis (the notion of logical form, arguments as propositions) and as the theory of evaluation (the notions of validity and soundness).

3 WHAT INFORMAL LOGIC IS NOT

Some readers will undoubtedly be familiar with Ryle's use of 'informal logic' in *Dilemmas* [Ryle, 1954, Chapter VIII], "Formal and Informal Logic". Ryle does not offer an explicit definition of informal logic, but develops the concept in contrast with formal logic, which he sees as tied up with the study of certain topic-neutral expressions (117). Ryle writes:

> ... there remains a very important way in which the adjective 'logical' is properly used to characterize both the inquiries which belong to Formal Logic and the inquires which belong to philosophy. [Note that the contrast is *not* between Formal and Informal Logic.] The Formal Logician really is working out the logic of *and, not, all, some*, etc., and the philosopher really is exploring the logic of the concepts of *pleasure, seeing, chance*, etc., even though the work of one is greatly unlike the work of the other in procedure and objectives. (119)

Thus, for Ryle, we might say that informal logic is the "logic" of substantive concepts, whereas formal logic is the logic of "topic-neutral expressions" (121). Ryle put the contrast this way:

> the handling of philosophical problems [cannot] be reduced to either the derivation or the application of theorems about logical constants. The philosopher is perforce doing what might be called 'Informal Logic', and the suggestion that his problems, his results or his procedures should or could be formalized is ... wildly astray ... (124)

While the entailment relationships of conjunction, disjunction and implication can be formalized, the implicative relationships of time and pleasure cannot. In this sense, informal logic becomes roughly equivalent to conceptual analysis, or to what Wittgenstein would call the "depth grammar" of an expression [Wittgenstein, 1953, p. 664]. Ryle's usage is not, then, particularly helpful in clarifying informal logic as we have come to understand that term.

A second view about informal logic is that it is simply and exclusively the study of the informal fallacies. Some might take this view from Carney and Scheer [1964] or from Kahane [1971], or, indeed, from the work of Woods and Walton on the informal fallacies during the decade 1972–1982 (collected in [Woods and Walton, 1989]), although they themselves did not use the term. We think, however, that although the study of fallacies does indeed constitute a part of informal logic's subject matter, it is only one part. Moreover, the study of fallacies does not belong exclusively to informal logic, since it forms part of the subject matter of other disciplines, such as rhetoric and speech communication, albeit from somewhat different perspectives.

A third view of informal logic is that it is whatever is left over in logic after the formal parts have been removed. We have suggested that this may have been what Carney and Scheer [1964] had in mind, and it is illustrated clearly by Copi's textbook, *Informal Logic* (1986), for that book consists entirely of the chapters from his earlier textbook, *Introduction to Logic* (1982 edition) that do not cover formal logic. There are problems with this characterization. On the one hand, in situations in which formalism is illuminating, informal logicians employ it (see [Woods and Walton, 1982]). On the other hand, the principal focus of informal logic is the logical evaluation of argumentation, which is not a "left-over", and over which in some cases informal analyses compete with formal analyses for the laurels of perspicuity and illumination.

A fourth view is that informal logic has the task of supplementing formal logic in the analysis and evaluation of reasoning in natural language. Such a view can be found in [Goldman, 1986] and [Woods, 1995]. On this view, where deductive implications occur in natural language argumentation, they can be formalized (sometimes with profit, sometimes not), but implications that are not entailments, and other aspects of argument analysis and evaluation, are not amenable to such formal treatment and so require the resources of an informal logic. Our own position is the reverse of this one: informal logic is the principal instrument for use in the logical analysis and evaluation of the reasoning found in natural language contexts, though occasionally the analysis of argumentation in natural language can profit from formal tools.

A fifth conception, that informal logic is applied epistemology, has found favour with authors such as McPeck [1981], Siegel [1988], and Weinstein [1994]. On this conception, informal logic is the study of the criteria of arguments that provide justification for beliefs or for knowledge claims. Since we do not think that it is the objective of all arguments or argumentation to support beliefs or knowledge claims, we do not think the norms for the epistemic function of arguments exhaust the subject matter of informal logic. Also, informal logic studies (non-deductive) inference relationships, a topic which we regard as belonging, if not to logic, then to pragmatics — not to epistemology, and if we are right, then informal logic cannot be reduced

to epistemology. (See [Scriven, 1987; Walton, 1997] for discussions of such inference relationships from a pragmatic perspective.) The overlap of informal logic and epistemology does not justify the reduction of the former to the latter, any more than the overlap between ethics and epistemology justifies the reduction of ethics to epistemology. In any case, the "reduction" of informal logic to epistemology risks skating in circles. For many, following Aristotle, formal deductive logic itself is embedded in epistemology in such a way that epistemology is applied logic!

Finally, we disagree with Fisher and Scriven [1997], who hold that informal logic is "the discipline which studies the practice of critical thinking and provides its intellectual spine" (76). Given our view that informal logic's defining subject-matter is arguments and argumentation, and given that critical thinking, in Fisher and Scriven's view, is "skilled and active interpretation and evaluation of observations and communications, information and argumentation" (21), we would assign to informal logic a narrower scope than they do.

4 THE AUTHORS' CONCEPTION OF INFORMAL LOGIC

Having said what informal logic is not, let us now focus on our understanding of what it is. We have characterized it as follows:

> *By 'informal logic,' we mean to designate a branch of logic whose task is to develop non-formal standards, criteria, and procedures for the analysis, interpretation, evaluation, critique and construction of argumentation in everyday discourse* [Johnson and Blair, 1977, p. 148]

We here use the term 'non-formal' as the contrast term to one of the three senses of 'formal' that Barth and Krabbe [1982, 14f] contend are to be found in logic, so we need to turn to their account to explain our meaning.

Barth and Krabbe use 'form$_1$' to identify the sense typified by the Platonic idea of form, where 'form' denotes the ultimate metaphysical unit. They contend that most traditional logic (for example, syllogistic logic) is formal in this sense. That is, syllogistic logic is a logic of terms where the terms could naturally be understood as place-holders for Platonic (or Aristotelian) forms. In this first sense of 'form,' almost all logic would be informal (that is to say, non-formal). Certainly neither propositional nor predicate logic can be construed as a term logic. Since we mean in part to distinguish informal logic from contemporary formal logic, our sense of 'informal' cannot usefully be understood as the contrast to a sense of 'formal' in terms of which *all modern logic* is informal.

By 'form$_2$,' Barth and Krabbe mean the form of sentences and statements as these are understood in modern symbolic logic. In this sense, one could

say that the syntax of the language to which a statement belongs is very precisely formulated or "formalized"; or that the validity concept is defined in terms of the logical form of the statements that make up the argument. In this sense of 'formal,' most modern and contemporary logic is "formal". That is, such logics are formal in the sense that they canonize the notion of logical form, and the notion of validity defined in terms of logical form so understood plays the central normative role.

It is in relation to this second sense of 'form' that informal logic is non-formal, because it abandons the notion of logical form as the key to understanding the structure of argument. It likewise abandons formal validity as constitutive for the purposes of the evaluation of arguments. When Govier speaks of "informal logic" (1987), it is this second sense of 'formal' that stands in the background. What we reject is the view that (with rare exceptions) the salient criteria for evaluating arguments are a function of their logical form.

According to what Barth and Krabbe distinguish as 'form$_3$,' informal logic is formal. They note a third sense in terms of which 'formal' refers to "procedures which are somehow regulated or regimented, which take place according to some set of rules". They say of their own position: "we do not defend formality$_3$ of all kinds and under all circumstances", but, "we defend the thesis that verbal dialectics must have a certain form (i.e., must proceed according to certain rules) in order that one can speak of the discussion as being won or lost" (19). Nothing in the informal logic enterprise stands opposed to the idea that argumentative discourse should be subject to norms, that is, subject to rules, criteria, standards or procedures, which can be specified and justified.

In sum, in the first sense of 'form,' all logic is informal; in the third sense of 'form,' all logic is formal. Only in the second sense of 'form' can the logic of argumentation, "informal logic", be distinguished from "formal deductive logic".

We do admit to an unwanted limitation in the definition of informal logic we stated above. It restricts informal logic to everyday discourse, and that seems both unnecessary and counterfactual. Historically, the focal point for informal logic has been what might be called natural-language argument, which may be divided into two sub-domains: *everyday discourse*, such as discussions of public affairs, for example newspaper editorials or arguments in legislatures (Govier calls it "ordinary argument"; others have called it "mundane argument") and what Weinstein [1990] called stylized arguments, namely, "argument within the various special disciplines" (121). We do not now believe that the domain of informal logic should be restricted by the subject matter of the arguments in question, or by the degree of theoretical specialization of their subject matter.

However, informal logic does focus on argument, and that leads to the next question: How does informal logic differ from formal logic? In our

view, the difference is partly one of subject matter. The social, communicative practice of argumentation can and should be distinguished from both inference and implication. In even the simplest such argument, the arguer proposes to the addressee that a set of propositions supports another, and thereby invites the addressee to infer the latter from the former.

The relation of support may, but need not be, one in which the supporting propositions purportedly deductively imply the supported proposition. Part, though not all, of the normative theory of argumentation has as its subject-matterthe answer to the question, "What sorts of support relationship are logically cogent?" We take it that one subset of logically cogent support relationships are ones in which the supporting propositions logically imply the supported proposition, and we take the subject matter of formal deductive logic to consist (*inter alia*) of the study of forms of entailment, that is, valid implication relationships among propositions or sentences. It will be relevant to whether the addressee ought to infer the supported proposition from those adduced as support for it to know that the latter logically imply the former, but that is neither a necessary nor a sufficient condition of a justified inference. The supporting relationship between premises and a conclusion need not be one of entailment. There can be justified inferences from premises to conclusions which are not deductively valid. In those cases, good premise-conclusion support relationships in arguments must meet other requirements, such as the relevance of the support, and, supposing its relevance, the sufficiency or completeness of the support. And there can be valid support relationships that are not cogent. For instance, question-begging arguments are valid, but fail to meet an epistemic or a dialectical norm (depending on the type of argument).

In addition, cogency includes standards that the grounds or premises of arguments must meet. Depending on the context, they must satisfy a requirement of truth, reasonableness or dialectical acceptability.

We take informal logic to be the study of all the conditions under which an addressee ought to infer the target proposition from those adduced by the arguer as support for it. Thus the subject matter of informal logic is different from that of formal logic. Informal logic has the task of developing norms for the evaluation of arguments; formal logic has the task of developing the norms for formally valid implication relations.

In saying that informal logic is about arguments and formal logic is not, we go against a use of 'argument' according to which logical proofs or demonstrations are counted as arguments, or sets of unrelated sentences or propositions are counted as arguments. For an example of the latter meaning, see Georgacarakos and Smith's textbook, *Elementary Formal Logic* (1979), where the authors say that according to their definition, the following is an argument (p. 94):

2 + 2 is 4.
Australia is in the Southern Hemisphere.
I'm hungry. (premises)
The moon is made of green cheese.
You owe me \$5. (conclusion)

In contrast, we mean by 'an argument as an entity' a set of reasons, more or less fully-developed, that is, has been, or might be, offered to another person or other persons with the intention of persuading them, or some audience, to modify their beliefs, attitudes or behaviour. And we mean by the activity of argument (or argumentation) various kinds of social interactions, normally predicated on disagreement, and usually including some element of the offering reasons to others for the purpose of persuading them, or some audience, to modify their beliefs, attitudes or behaviour.

5 OTHER PERSPECTIVES ON INFORMAL LOGIC

The above constitutes our account of informal logic, but for the record and because the term has not yet achieved an accepted general usage, we need to discuss also some of other accounts. We have selected Govier [1987] and Walton [?], for they are not only two of the most prolific theorists in the field at the moment, but also (deservedly) among those most widely respected.

Govier. For Govier [1987], informal logic denotes the art of argument evaluation, a task she insists is non-formal in character:

> Informal logicians, contesting the drive for complete rigour and purity among formalists, contend that the appraisal of natural arguments requires something other than translation into a technical formal language and application of formal rules to test validity. However, it is only in the stricter sense of 'formal' that informal logicians are committed to the view that a theory of natural argument would not be a formal theory. (14)

Govier then adumbrates a distinction between different senses of 'formal' similar to the one we introduced above. For her, informal logic is the logic of "natural" arguments:

> To speak of informal logic is not to contradict oneself but to acknowledge what should be obvious: that the understanding of natural arguments requires substantive knowledge and insights not captured in the rules of axiomatized systems. The informal fallacies, historically a central topic for informal logic, involve mistakes in reasoning which are relatively common, but neither formally nor informally characterizable in any useful way. The fact that an account of informal logic makes it out to be just that does not show that it is imprecise or lacking in rigour. (204)

Thus for Govier informal logic is the logic which helps evaluate what she calls natural arguments — a process requiring substantive knowledge and insights not provided by formal logic. Noteworthy as well is Govier's connection of informal logic with the informal fallacies.

Walton. Walton's [1998c] paper deals explicitly with the issue of the relationship between formal and informal logic:

> Formal logic has to do with the forms of argument (syntax) and truth values (semantics)... Informal logic (or more broadly, argumentation, as a field) has to do with the uses of argumentation in a context of dialogue, an essentially pragmatic undertaking. (418–19)

Thus for Walton, informal logic is about the pragmatic aspects of argument, whereas formal logic is about the semantic and syntactic aspects of it. More recently (1998), Walton has described the difference between formal and informal logic as follows:

> Formal logic evaluates arguments as correct or incorrect insofar as a given argument is an instance of certain forms of argument that are recognized as valid. ... There is some dispute about exactly how the forms should be applied to actual cases [Massey, 1975]. But formal logic does at least have the decisive asset that the form functions as an objective standard for evaluating the worth of an argument in some clear respect.
>
> Informal logic has long suffered from the lack of a clear standard as an objective basis for evaluating arguments. This lack of a relatively simple criterion may be partly a reflection of the fact that informal logic evaluates how an argument is used in a given context — a more practical and applied task that tries to judge real cases of arguments as they occur in a natural-language setting. Such an applied task of evaluation would understandably involve more interpretation of what the argument is supposed to be within a common framework of understanding common to the arguer and the audience or respondent.
>
> ... the field of informal logic is at a crisis point. How, generally, can we distinguish between the fallacious from [sic] the non-fallacious instances of these arguments? (9–10)

The gist of Walton's view is that informal logic is necessarily pragmatic in character, given its subject-matter, namely, the evaluation of real arguments in their actual settings. Formal logic may be applied to this task, although controversially. However, it identifies only one aspect of such arguments. Since Walton believes informal logic faces a theoretical crisis — one which

he presents his theory to resolve — he believes that the "aspect" of an argument whose worth is clearly evaluated by the application of formal logic is very far from all that is significant to note in the evaluation of arguments. He does not regard the two areas of logic as in competition, but rather as complementary.

6 SUMMATION

In this section, we have discussed the nature of informal logic. We looked at its historical antecedents, we then discussed what informal logic is not, and we followed that with our own treatment of informal logic and most recently have featured other conceptions of informal logic.

IV. Issues and Approaches

1 INTRODUCTION

Having discussed the meaning and the origin of informal logic, we turn in this section of the chapter to a discussion of the various issues that informal logic now treats and then to an account of the approaches that having been developed by its practitioners.

2 THE MAJOR ISSUES IN THE THEORY OF ARGUMENT

What are currently the main issues in informal logical belong to what we call the theory of argument, which we now subdivide into the theory of analysis and the theory of evaluation.

In the theory of analysis, the major issues are:

- how to understand the nature of argument in general: some stress the varied purposes that argument serves, others stress one or another function in particular;
- the elements that constitute an argument: for many researchers, the traditional way of breaking an argument down — premise, conclusion and assumptions — is adequate; others, following Toulmin, believe that the traditional approach is too closely wed to mistaken assumptions and needs to be replaced by Toulmin's model of grounds, warrant (plus backing) and claim (1958);
- how best to understand the structure of argument: If one gives up the traditional formal logical approaches in terms of logical form, then how is one to understand the structure of argument? There has been

much work on different ways support for a conclusion might be ordered, starting with the "linked vs. convergent" distinction, but going beyond it;

- the general issues of the interpretation of argumentative texts: this work includes the fundamental hermeneutic principle, charity, which has been debated and understood in various ways;
- the question of how, based on these understandings, to formulate tools and procedures for the interpretation of particular arguments in their contexts: this topic includes how best to display the structure of the argument, as well as methods for dealing with other problems in the task of argument interpretation.

In the theory of evaluation, the main issue concerns the normative question: "By what standards and according to what theory should arguments be evaluated?" Specifically, "What norms apply to the transfer of support from the premises to the conclusion?"

A related topic is the large question of the role of fallacies in the logic and the theory of argument. Here a distinction is needed between (a) fallacy theory as whole, since the very concept of fallacy has come under attack, and (b) the analyses of the individual fallacies.

2.1 Argument analysis

The shift of pedagogical attention from invented examples of arguments and old historical saws to actual arguments in their original form and settings had great practical and theoretical consequences for the development of informal logic. These "natural" (for naturally occurring) arguments are ill-organized, incompletely stated, meandering and multi-functional. As we have argued above, the teacher standing before a class of critical, questioning students could not readily pretend that the simple forms of the propositional or predicate calculus neatly map the formidable complexity of actual arguments. The experience tended to force the teacher to become a seat-of-the-pants theorist. Even so, informal logicians have tended to minimize the tangle of theoretical issues raised in the attempt to explain how we manage to interpret such discourse. they have left that task to linguists, speech communication theorists, and philosophers of language. The doctrine on interpretation in informal logic textbooks has been *ad hoc* rather than theoretically motivated, based on generalizations from experience rather than on systematic study or tested theory (Fogelin [1978], which introduces speech-act theory and Gricean implicature, is arguably an exception). Where informal logicians focussed their attention instead was the issue of how to lay out the sentences playing argumentative roles in such discourse in a perspicuous way, for the purpose of evaluating the argument.

Argument structure

As we have noted, Toulmin's [1958] model initially had less influence among informal logicians than it had in argumentation theory generally, and in speech communication in particular. Beardsley's diagrams [1955; 1966] — an approach to displaying the structure of the support for the conclusion of arguments, including the premise-conclusion relations of "internal" or premise-supporting arguments — have been more influential. Beardsley distinguished between "convergent", "divergent" and "serial" lines of support. Support in an argument is convergent, if "several independent reasons support the same conclusion"; divergent, if "the same reason supports several conclusions"; and serial if it "contains a statement that is both a conclusion and a reason for a further conclusion" [1955, 19]. Thomas later added the concept of "linked" support: "when a step of reasoning involves the logical combination of two or more reasons" [1973, 52]. He showed such linked reasons typographically as sentences side-by-side on the page with a "+" sign standing between them. Beardsley's diagrams showed each claim supported separately by one or more reasons. Thomas's innovation permitted distinguishing between two or more reasons supporting a conclusion "in a united or combined way", and two or more reasons each supporting the conclusion "completely independently and separately" (54). (This is the distinction between, "The accused is guilty because the murderer was left-handed and so is the accused", and "The accused is guilty because he confessed, and anyway we found his DNA in the blood on the murder weapon".)

Beardsley's [1955] and Thomas's [1973] diagrams contained each premise written out in full, with arrows pointing from each reason to the claim it supports, but in another textbook (1966: *Thinking Straight, Principles of Reasoning for Readers and Writers*), Beardsley assigned a number to each proposition in the texts he constructed as examples (bracketing each proposition and writing its number in a circle right in the text), and then drew diagrams with each arrow pointing from one number to another, with the number at the arrowhead representing the supported claim and the number at the other end of the arrow representing the supporting reason.

Following on these developments, Scriven (1976) added some new features. He used the same bracketing and number method as Beardsley (1950b), and combined a quasi-arrow version of Beardsley's diagrams (a line between numbers, but without an arrowhead pointing to the conclusion) with Thomas's "+" sign indicating premises linked in their support for their conclusion. Scriven introduced the convention of using a minus sign ("−") to symbolize "negative" support, as when the arguer concedes an objection to the conclusion, thus making it possible to use the diagram convention for "balance of considerations" arguments. These are arguments that balance reasons favouring a conclusion against reasons opposing it and produce an "all things considered" conclusion. Scriven's second innovation

was to use a letter of the alphabet (as distinct from the arabic numerals used to number assertions) to identify each reconstructed "unstated assumption" of the reasoning, which he suggests that the person doing the analysis should write out and so label. The letters representing these unstated assumptions are to be placed in the diagram at the appropriate place to portray the structure of the argument so as to show their role in joining with stated premises to support the inference from the stated premises to the conclusion (43).

Johnson and Blair [1977] followed Scriven's conventions for the fine-grained details of arguments, but in (1983 and 1993/1994) they introduced a modification in order to diagram the "macrostructure" of extended arguments, that is, argumentative texts in which the author tries to make a more or less fully-developed case for the claim in question. In such texts there are normally not only more than one line of argument in support of the conclusion, but also lines of argument that support the conclusion only indirectly, by aiming at refuting objections to the arguments for the conclusion or objections to the conclusion itself. Johnson and Blair used letters to represent the paragraphs of the text, and placed all the letters representing the paragraphs developing each directly-supporting line of argument inside one circle, with an arrows to the conclusion showing the direction of support. They use dotted lines with arrowheads pointing away from the conclusion to represent objections mentioned in the argument, and solid lines with arrowheads pointing back to the conclusion to represent the fact that a line of argument is presented as an answer to such an objection instead of as direct support for the conclusion. Later texts have modified and amplified these basic approaches to the task of representing the structure of arguments.

Recently, the conceptions of argument structure underlying such diagramming have come under critical scrutiny. For example, there is an unresolved ambiguity as to whether each unit in the diagram is a distinct premise or an argument as such — a set of premises. The so-called linked structure seems to capture sets of premises working jointly, whereas the so-called convergent structure seems to capture different lines of argument. Two monographs discussed argument-structuring and diagramming at some length. Freeman [1991] combined modifications to the Toulmin model with a dialectical analysis. So, for example, he sees a linked argument as one in which a relevance-warrant is added to a statement of data in answer to the question why the data was relevant to the conclusion (p. 94), whereas a convergent argument is one in which the questioner has asked for *another* reason (p. 95). Snoeck Henkemans (1992), applying the pragma-dialectic argumentation theory of van Eemeren and Grootendorst (1983), argued that conceiving arguments as dialogical discussions aimed at dispute resolution permits the removal of confusions about the linked-convergent distinction. According to her theory, the structure of an argument should be understood as a function of the dialectical roles of its com-

ponent propositions. Both Freeman and Snoeck Henkmans should be consulted for their extensive, theoretically-motivated elaborations of argument diagramming. Other important discussions of the "linked-cnvergent" distinction are found in articles by [Yanal, 1991; Conway, 1991; Vorobej, 1995a; Vorobej, 1995b].

Missing Premises

There has been considerable discussion in the informal logic literature of questions surrounding the problem of missing premises (see Hansen's [1990] index entry for a list of articles). Attention to the structure of actual examples of arguments has invariably been accompanied by the attempt to account for what are variously called "assumptions","missing premises", "hidden premises", "unexpressed premises", "tacit premises" or "unstated premises". If various indicators in the text make it clear that an arguer clearly intends one statement to support another, but it does not, there is a presumption that the arguer expected the listener or reader to understood as read some additional proposition(s) that, when added to the initial statement, enable the entire set to be relevant, or to provide stronger support than the expressed propositions supply by themselves. Aristotle's enthymemes are an instance. But how shall the "missing premise" (as we will call it) be formulated? Which argument should be reconstructed– the speaker's, the strongest possible (given what's stated), or some other? Scriven, for example, distinguished between the "arguer's assumptions" (what the arguer consciously assumed or would grant), the "minimal assumptions of the argument" (what is logically speaking necessary to get from the premises to the conclusion) and the "optional assumptions" (claims which are logically adequate and well supported) [1976, 43]. Should the argument be reconstructed so the premises logically imply the conclusion (so that the logical form of the resulting argument is deductively valid), or so that they provide strong support, or merely so that they are relevant, or something else? Should the missing premises supplied be (likely to be) believed by the arguer, known to the arguer, what the arguer (likely) expected the listener to believe, or something else? In an early influential article, Ennis [1982] distinguished between used assumptions and needed assumptions. The former determine the missing premise considering the argument from the point of view of the intent of the arguer, the latter determine it considering the point of view of the grounds required to support the conclusion. Van Eemeren and Grootendorst [1992] propose that logical and pragmatic norms are required in the reconstruction of unexpressed premises. An unexpressed premise is required when the argument would otherwise be invalid (not necessarily deductively invalid — invalid "according to whatever validity criterion is preferred" (60). But the "logical minimum" (the associated conditional of the argument: a conditional formed by the stated premises

as the antecedent and the conclusion as the consequent), which turns the argument into a case of modus ponens, is ruled out as misrepresenting the speaker's argument. Instead, they propose that the unexpressed premise be the "pragmatic optimum": the premise that makes the argument valid and also prevents a misrepresentation of the speaker's actual position and any other rule of communication (64).

The principle of charity

Any discussion of the issue of missing premises is bound to bring to the fore the issue of how best to interpret argumentation in natural language, where there are often indeterminacies and various possible ways of handling them. The principle of charity is a principle of interpretation which in its most general formulation could be put this way: "When interpreting a text, make the best possible sense of it". This would mean, for instance, that given a choice between two interpretations, the principle of charity advises us to choose the one which puts the text in the best light, given its purposes. Thus if we have a choice between interpreting a text as a fallacious argument and interpreting it as ironic discourse, then charity requires that we adopt the latter interpretation.

The principle of charity has its origins in hermeutic inquiries about how best to interpret texts (see [Govier, 1987]; for a useful discussion of its recent history in semantical issues in philosophy, see [Stein, 1996, Chapter 4]). It should be noted that there is some disagreement about how to formulate the principle as it is applied in the theory of argument (see [Johnson, 1981; Govier, 1987].

The principle of charity applies at a number of levels. It applies in deciding how to interpret a text; it applies when we are deciding which of several candidates for the missing premises to assign to the argument, and it also applies in attempting to decide both the structure and the content of the argument. However the principle is to be formulated, what we are concerned with here is its role in informal logic, particularly in connection with the theory of analysis.

It is instructive to note that no counterpart to this problem seems to have existed for formal deductive logic. To the extent that formal logic textbooks deal with arguments in natural language, the arguments tend to be either invented examples tailored to illustrate context-independent logical relations, or arguments in use selected because they transparently exhibit context-independent logical structures. Hence the need for interpretation is non-existent or minimized. On the other hand, informal logic considers any and all arguments in use, with antecedently-formulated structures supposedly empirically-grounded, and subject to correction, modification, or rejection by the empirical data. Hence the interpretation of such arguments is context-dependent to the maximum possible degree. For further

references, see [Hansen, 1990, p. 181]. The principle of charity, its proper formulation and justification, continues to be a topic of interest to informal logicians.

2.2 Theories of evaluation

Perhaps the central issue for informal logic is the appropriate theory of argument evaluation or criticism. Having exposed and illuminated the structure of the argument and given as plausible a reconstruction as we can in accordance with the principle of charity, we face the question: How good is this argument? The informal logician must be able to answer this question for arguments for which the criterion of "soundness" (true premises and deductively valid relation between premises and conclusion) is incomplete, unilluminating, or inappropriate. How have informal logicians answered this question?

Their answers make up what we have termed the different approaches to the theory of evaluation within informal logic, which we will now discuss.

2.2.1 Fallacy theory as a theory of evaluation

Given the way that informal logic has developed in close association with the study of fallacy, it is not surprising that fallacy theory has presented itself as the dominant theory of evaluation in informal logic. A good argument will be one that is free of fallacy, and the presence of a fallacy is a prima facie weakness, if not a fatal flaw, in the argument. The problem with this intuition is that there is little consensus about the correct general theory of fallacy or about how to use fallacies as tools for argument criticism. Indeed, some hold that no fallacy theory can be viable [Finocchiaro, 1981; Massey, 1981]. (See our discussion of their arguments below). We proceed now to an outline of some of the contending theories of fallacy.

In his seminal work, *Fallacies*, C.L. Hamblin (1970) claimed that "the standard treatment" of fallacies in the textbooks is to be described as follows:

> What we find in most cases, I think it should be admitted, is as debased, worn-out and dogmatic a treatment as could be imagined–incredibly tradition-bound, yet lacking in logical and historical sense alike, and almost without connection to anything else in modern Logic at all. (12)

Hamblin called for a concerted effort to revitalize this impoverished corner of logic, and beginning in the early 1970s, John Woods and Douglas N. Walton took up Hamblin's implied challenge, and published together over the course of the next ten years 19 papers, mostly on individual fallacies but some on fallacy theory in general (collected in [Woods and Walton, 1989]).

Their approach was *ad hoc* in the sense that they worked from no global theory of fallacy, but it was very much in Hamblin's spirit in that they sought historical roots, provided thorough original analyses of the fallacies they studied, and made ample use of the tools of modern logic when and as those proved illuminating. We might call their approach to fallacy the "analytical-logical" approach.

Another approach that has been followed by several textbook authors makes fallacies out to be patterns of violation of a set of criteria of argument cogency. The criteria in question emerged from consideration of the various types of fallacy. There are fallacies of relevance (such as *ad hominem*), fallacies of sufficiency (such as hasty conclusion) and fallacies of acceptability (such as inconsistency). According to this theory of fallacy, the premises of a cogent argument must be acceptable (vs. true) and provide relevant and sufficient support for the conclusion (vs. deductively imply it). A fallacy is then a violation of one or more of the criteria of acceptability, relevance and sufficiency. This account first appeared in [Johnson and Blair, 1977] and later in [Govier, 1985; Damer, 1987; Freeman, 1988; Little *et al.*, 1989; Barry, 1992] and [Seech, 1993]. We might call this the RSA approach. A few words about its elements follow.

Relevance. There is no widely accepted account of probative relevance–the relevance of premises for conclusions. This is unsurprising, since relevance has received little attention from informal logicians until recently. Grice [1975] makes relevance one of the criteria for conversational implicature, but his account assumes an understanding of relevance and provides no theoretical enlightenment of that standard. Sperber and Wilson [1986] have the most definitive account of relevance to date, but it does not settle the nature of probative relevance. A conference on relevance in 1991 at McMaster University revealed no consensus (see the papers published in *Argumentation* 6:2, 1992). There Hitchcock [1992] argued that relevance is a triadic relationship between an item, an outcome or goal and a situation, and further for the need to distinguish epistemic from causal relevance. Woods [1992] has argued that relevance cannot be understood as a relationship between ideas or propositions but rather must be understood in terms of a broader unit–the agenda.

Second, it is understood that sufficiency presupposes relevance [Biro, 1992]. If the premises of an argument do provide sufficient support for its conclusion, they are ipso facto relevant. But some [Johnson and Blair, 1994a; Govier, 1987] see relevance as a distinct requirement because there are some problems with arguments that are specifically relevance failures, e.g. certain forms of *ignoratio elenchi*, the *ad hominem*, and red herring.

Third, some cases of irrelevance in arguments might be seen as violations of procedural rules, such as "restrict your criticism to the position your opponent holds", and not logical inference rules (see [Eemeren and Grootendorst, 1992]), for example, straw person.

Sufficiency. Of the three criteria, this one has so far attracted the least amount of sustained discussion. No one seriously believes that anything like an algorithm for sufficient evidence will ever be available. The question whether the premises provide sufficient evidence for the conclusion has three dimensions. First, are all the appropriate types of evidence presented? Second, within each type, is enough evidence of that kind provided? Third (what might be termed dialectical sufficiency), does the arguer respond appropriately to known or plausible objections or incompatible evidence? (See [Blair, 1992; Johnson and Blair, 1994a; Johnson and Blair, 1994b], and [Johnson, 1996, pp. 264–165].

Acceptability. The criterion of acceptability is the informal logician's counterpart of the truth-requirement of the traditional doctrine of soundness associated with formal deductive logic. When informal logicians rejected the latter ideal as an adequate theory of argument evaluation, they likewise tended to reject the criterion that the premises of a good argument must be true. Influential in this respect was Hamblin [1970], who argued that truth was an inappropriate criterion for arguments with a given audience, context, and occasion, being both too weak (we want premises that are not only true, but known to be true) and too strong (in many contexts in which argument is appropriate, truth is a rare commodity, for example in courts of law). Hamblin's complaint was not unlike that of the deconstructionists: the idea of truth presupposes an impossible God's-eye position from which to view matters. Hamblin urged that we accept instead the requirement that the addressee be ready or willing to commit to the premises — that is, that they should be accepted by the addressee. Blair and Johnson [1987] proposed a normative version of Hamblin's "accepted" requirement, namely, that premises be worthy of acceptance, and attempted a specification of that norm. Many informal logicians have adopted acceptability as a norm of premise-adequacy, but Johnson [1990b] has had second thoughts about it, arguing that on any construal acceptability is problematic, and in any case it may be too weak a criterion. More recently, some informal logicians have argued that both truth and acceptability should be a requirements for premise-adequacy (see [Johnson, 1996; Johnson, 1999; Allen, 1998]). This approach might be called the informal logic approach to fallacy.

An entirely different and novel theory of fallacy was proposed by E.M. Barth [1982], who, citing the influences of Ness, Crawshay-Williams, Hamblin, Lorenz and Lorenzen, suggested that a fallacious argument be understood as "an argument not generated (at that stage of the discussion) by the system of dialectical rules" governing it (Barth referred also to her work with Krabbe [1978; 1982]). She suggested that:

> the study of fallacies can then move from the stage of a catchword classification that we know today to an investigation of the

> relation of a given argument (i) to the dialogical situation at the given stage of the discussion and (ii) to the dialectical rules (of a given system) that pertain to exactly that dialogical situation. (160)

This might be called the dialogical approach to fallacy.

The dialogical approach was developed in a distinctive way by the discourse analysts van Eemeren and Grootendorst [1983; 1992]. They contend that argumentation presupposes (or may be modelled by) an ideal model of a discussion between disagreeing interlocutors which is aimed at rationally resolving their dispute. The rationality of such an argumentative discussion is a function of constitutive rules which regulate it, rules which derive their rationale from the purposes of the discussion. The classical informal fallacies, van Eemeren and Grootendorst maintain, are best understood as particular types of violations of the rules that are constitutive of rational discussions aimed at resolving disagreements. They not only make this proposal in general terms, but they also attempt to show in detail how each of the traditional fallacies violates one or another of ten rules of critical discussions. This approach has been called the "pragma-dialectical" theory of fallacy.

Walton's work on fallacy

Over the last fifteen or so years, Douglas Walton has been working out a distinctive, theoretically unified theory of fallacy, publishing elements and successive iterations of it in a series of monographs (1984 *Logical Dialogue Games and Fallacies*; 1987 *Informal Fallacies*; 1989a *Informal Logic*; 1989b *Question-Begging Argumentation*; 1990a *Begging the Question*; 1992a *The Place of Emotion in Argument*; 1992b *Plausible Argument in Everyday Conversation*; 1992c *Slippery Slope Arguments*; 1995 *A Pragmatic Theory of Fallacy*; 1996a *Argument Schemes for Presumptive Reasoning*; 1996b *Arguments from Ignorance*; 1997 *Appeal to Pity*; 1998a *One-Sided Arguments: A Dialectical Analysis of Bias*; 1998b *Ad Hominem Arguments*). Walton accepts a traditional general account of fallacy: "A fallacious argument involves an error of reasoning but also a misuse of reasoning, in many cases, as a sophistical tactic to try unfairly to get the best of a speech partner in a dialogue by deception" [1998b, 292], but he has a distinctive view as to the nature of such mistakes in reasoning. Walton's theory shows the influence of the dialogue-game and pragma-dialectical approach in that he regards argumentation as modelled by dialogue, and he takes its norms to be importantly pragmatic as well as syntactic and semantical. But he parts company with van Eemeren and Grootendorst when he argues that there are many different types of dialogue occurring typically in argumentation, not just disagreement-resolving (or "persuasion") dialogues, and he is not

concerned to try to formulate a list of the rules that any rational argumentative dialogue must obey. When engaged in one or another of the various types of dialogue, often moving from one type to another during the same conversation, we employ a wide variety of types of patterns of argument or "schemes", many of which have numerous species (for instance, [Walton, 1998b] distinguishes 14 species of *ad hominem* argument schemes, and five sub-species of one of these). A fallacy, in Walton's more specific account, is a fatal failing in the use of an argument scheme, either internally, by way of a failure to fulfil its particular conditions, or externally, by way of the use of an inappropriate scheme — an illicit dialogue shift, in which a scheme appropriate to one kind of dialogue is improperly deployed in a different type of dialogue. These features account for both the erroneousness and the deceptiveness of fallacy. Walton calls his approach a "pragmatic" theory of fallacy.

The last approach to fallacy we will mention has emerged from Hintikka's inquiry-based theory of reasoning. Hintikka and Bachman [1991] propose that reasoning be understood as rational inquiry, which can be modelled by interrogative or questioning games. They distinguish between definitory and strategic rules, and between logical inference rules and interrogative rules, and thus produce a "fourfold classification of fallacies: 1. Definitory logical inference fallacies. 2. Strategic logical inference fallacies. 3. Definitory interrogative fallacies. 4. Strategic interrogative fallacies" (412). Since a strategy involves both logical and interrogative moves, strategic fallacies may be a combination of 2 and 4 (413). We might call this the interrogative approach to fallacies.

Thus we have discussed here briefly no less than 5 different approaches to fallacy: the informal logic approach, the dialogical approach, the pragma-dialectical approach, the pragmnatic approach and the interrogative approach.

2.2.2 The natural-language approach to argument analysis

There are many theorists who approach the task of argument analysis eschewing all or most talk of fallacies. A good example of this is found in [Scriven, 1976]. Scriven opposes using fallacies as the main tool of argument criticism. He argues that doing so requires building into the identification process all the skills needed for analysis without the taxonomy, and he claims that turns fallacy criticism into a formal approach with a tricky diagnostic step (1976, xvi). He thinks there already exists in natural language a sufficiently rich vocabulary suitable for critical purposes, for example: 'reason,' 'evidence,' 'conclusion,' 'thesis,' 'relevant,' 'sufficient,' 'inconsistent,' 'implication,' 'presupposition,' 'objection,' 'assumption,' 'ambiguous,' and 'vague,' among others. Scriven's method of evaluation is to raise the questions embedded in steps 5-7 of the seven-step method cited above: "(5)

Criticism of (a) The premises (given and "missing"), (b) The inferences, (6) Introduction of other relevant arguments, (7) Overall evaluation of this argument in the light of (1) through (6)".

For criticism of the premises, Scriven adopted a criterion of "reliability" (43). In criticizing the "inferences" — the premise-conclusion connection — he asks the critic to inquire whether, given the truth of the premise(s), the conclusion is true or likely to be true. He thus broadened the conception of validity beyond syntactic and semantic entailment in ways not unlike Thomas (1973). The test of an inference (as well as of a generalization, a definition or an interpretation or analysis) is whether it withstands counterexamples. He requires the arguer to look for other arguments that bear on the issue, both those pointing to a different conclusion and also other arguments favouring the stated conclusion. Thus one must go outside the frame of the argument in order to evaluate it. And the critic must in the end make an overall judgement. (See [Hitchcock, 1983], for an application of Scriven's approach.)

2.2.3 Field theories

The last approach we shall mention is the field theory. Toulmin and others [McPeck, 1981; Weinstein, 1994] take the view that standards of good argument are specific to each field or discipline (see [Toulmin *et al.*, 1978]). We have seen that McPeck [1981] and Weinstein [1990] think the important standards for an argument are furnished by the epistemology of its field. The middle ground position between the "all standards are field dependent" view and the "important standards are field invariant" view (the latter includes some-invariant standards as well as field-dependent standards. For a more thorough treatment of the issues that surround field theories, see [Johnson, 2000, pp. 291–309]).

In summary, at this point, it seems clear that there continue to be divisions among theorists in informal logic about the kind of normative theory which should be used in the evaluation of argumentation.

3 FALLACIES AND FALLACY THEORY

We have just discussed the main approaches to fallacies in broad outline, but given their centrality (their long association with informal logic) and controversiality as instruments of criticism and evaluation, we believe that more needs to be said about them.

3.1 The viability of fallacy theory

There has been an enduring interest in the question: Is the concept of fallacy viable? In other words, is there any legitimate subject-matter for a theory of fallacy? The viability of fallacy theory has its opponents and its

defenders. The latter subdivide into those who believe a unified theory of the fallacies is necessary, and those who do not.

Hamblin [1970] first raised the question of the viability of fallacy theory in general by showing that the treatment of fallacies found in the textbooks were often conflicting, confused, and on the whole completely inadequate. Hamblin himself was not a sceptic about fallacy theory and seemed to hold out hope for a theoretical basis for fallacies in formal dialectic:

> ...if the scattered survival of discussions of fallacies requires a justification other than in terms of entertainment value, this [formal dialectic] is what it must be. We need to see our reasoning in the kind of context within which, alone, these faults are possible. (1970, 255)

But a decade later two influential articles, building on Hamblin's critique, challenged the viability of fallacy theory. In the first, Finocchiaro [1981] not only criticized the accounts of fallacies found in textbooks (in the spirit of Hamblin but looking at a different selection), but also suggested that the very charge of fallacy is frequently a creation of the critic. In the second, Massey [1981] also criticized textbook accounts of fallacy and went on to argue that there cannot, in principle, be a logic of fallacies. His argument is based on his independently-defended position that there can be no theory of invalidity. If there can be no theory of invalidity, and if Hamblin is right that a fallacious argument is one that appears to be valid but is not, then it follows, Massey argued, that there cannot be a theory of fallacy.

In defence of fallacy theory, Govier [1982] argued that the challenges raised by both Finocchiaro and Massey make important but undefended assumptions regarding the theory of argument. Johnson [1987] argued against some of the traditional criticisms of fallacy theory, such as that the ways of going wrong are infinite and hence that there cannot be any theory of fallacy, nor any reliable classification. Johnson [1989] also contended that Massey's position rests on a questionable conception of fallacy — -i.e., that a fallacy is properly thought of along the lines proposed by Hamblin as an argument that appears to be valid but is not. Johnson argued that this conception of fallacy is deficient. Moreover, Massey makes assumptions about what an adequate theory must look like that are biased against informal logic.

Woods [1992] offered a defence of fallacy theory. He argued that he classical "gang of eighteen"(27) informal fallacies "caricature epistemic (and sometimes strategic) misperformance of [rational] skills necessary for our survival" (33). Not only are the skills in question important, but many of the misperformances appear to be imbedded in human nature. In addition, Woods contended that the correct analysis of many of the fallacies turn on deep and vexing philosophical issues, with the result that "fallacy theory is continuous with philosophy itself, and yet it branches into all the disciplines" (33).

At this point in the debate, the objections directed at fallacy theory, however serious, have not been sufficiently persuasive either to cause the abandonment of the concept of fallacy or to dissuade theorists from the attempt to give better accounts of fallacy theory in general and of the individual fallacies in particular.

Among those sympathetic to fallacy theory the main bone of contention, apart from disputes over the best analyses of individual fallacies, is between those who argue for a unified theory of fallacy, one that is not tied exclusively to logic, and those who believe that a unified theory might be impossible, and that proof-theoretic treatment is a powerfully illuminating tool worth trying for. (Clearly the second pair might be compatible.) Van Eemeren and Grootendorst maintain that "ideally, one unified theory that is capable of dealing with all the different phenomena, is to be preferred" and "it is important not to exaggerate the role of logic" [1992, 103]. In their joint work, Woods and Walton were content to analyse fallacies one at a time, trying to use a variety of logics to "provide solid generalizations" [1989, xix], and Woods has argued that any integrated theory should reflect the differences in the rational survival skills the fallacies caricature. He finds the task of producing a unified theory that respects such non-trivial levels of complexity to be "daunting", and concludes: "Better ... the present course of tolerant methodological pluralism . . ". [1992, 43]. As noted above, more recently Walton (e.g., [Walton, 1996a]) has been developing a unified theory of fallacy, albeit one that is more complex than the pragma-dialectical conception. One can expect to see more of this debate.

A different controversy exists over the utility of using the informal fallacies as a pedagogical tool in teaching students how to analyze and evaluate arguments. Scriven [1976] early on voiced scepticism, holding that "the fallacies generally turn out not to be fallacies — -unless only builds into the identification process, and hence into the labels, all the skills needed for analysis without the taxonomy of fallacies. In that case one has made it a formal approach, and the encoding (i.e., diagnosing) step has become the tricky one" (p. xvi). More recently, Hitchcock [1995] and Blair [1995] address the question of whether fallacies have a place in teaching reasoning skills. Hitchcock argues that the many of the traditional fallacies are not common; most are instances of patterns of reasoning that are not always fallacies, so time is better spent teaching the canons of good reasoning; students learning fallacies tend to become over-critical; and the fallacies contribute nothing to learning how to construct good arguments. Blair, on the contrary, thinks that with care the dangers Hitchcock mentions can be avoided. Fallacies provide good-sized learning chunks and learning them requires learning good argument analysis, the standards of good argument, evaluation skills and critical discrimination — -provided the analyses of fallacies are current and their treatment is not rushed.

3.2 Studies of the individual fallacies

Hamblin's criticisms struck a nerve. As we have noted, they inspired a large body of work by Woods and Walton, although recently there has come a call for a re-evaluation of Hamblin's critique (see Johnson [1990a; 1990b]). Where do matters stand now with respect to accounts of the individual fallacies? Are there now definitive analyses of the traditional fallacies? Are there new fallacies worthy of inclusion in the repertoire? There can be little question that the current analyses of most fallacies are more clear and precise than ever before in the history of the subject.

Following on his joint work with Woods in the 1970s, Walton has set the standard for tighter and more-finely focussed book length studies of fallacies in the last decade, as noted above, covering *ad hominem*, begging the question, appeal to pity, slippery slope, and appeal to popularity. Walton views fallacy in the context of argumentation, and his work shows the benefits of the rhetorical concern for context and audience (see [Walton, 1992a]). Walton's work on fallacies deserves both commendation and more detailed commentary than we can devote to it here. There have also appeared many more article-length studies of the individual fallacies, (for example [Brinton, 1985] and [Wreen, 1987]). For a useful bibliography of recent work on fallacy and particular fallacies, the reader should consult H. V. Hansen, "An Informal Logic Bibliography" (1990), which listed, since Hamblin's book appeared, some twenty-two articles on *ad hominem* and *tu quoque*, ten on *ad verecundiam*, thirteen on other "ad" fallacies, and thirty-one on all other fallacies. Hansen's more recent bibliography in Hansen and Pinto [1995] lists by our count 158 journal articles on particular fallacies, in addition to monographs, and to articles on fallacy theory in general.

By way of summing up this discussion of fallacy theory, it seems safe to say that in spite of challenges and criticisms, the topic of fallacy has remained a significant focus for research.

V. Future Research

1 INTRODUCTION

If our account of informal logic has not made it clear already, let us be explicit about the list of items that belong on the agenda for future research: everything is up for grabs! There is virtually no area of the field that is widely agreed to be settled, its theoretical house in order, with one doctrine that is (for the time being) beyond dispute. That should not be taken to imply that enormous strides have not been taken since the First International Symposium on Informal Logic in 1978, the very first scholarly meeting to discuss informal logic. On the contrary, the past two decades

have witnessed both a burgeoning of research and notable theoretical developments on many fronts. However, matters need to be put into perspective. Twenty years is a short time in the history of ideas. The revival of modern (deductive) logic began in the mid-19th century and it took over fifty years to see widely-accepted results, with developments now ongoing for 150 years. Informal logic has not yet reached that state of development.

More helpful than the comment that "nothing is settled" would be a list of the moot questions, and we offer such a list below. Any attempt to organize these questions into some sort of coherent order requires making even more contentious judgements than does selecting and describing the questions in the first place. Thus we acknowledge that, our best efforts to the contrary, the following list cannot avoid being selected, and organized and formulated, in ways that will reveal our own perspective and biases.

2 A DOZEN OPEN QUESTIONS IN INFORMAL LOGIC

(1) The nature of argument and argumentation

Does argumentation have essential properties such as dialogue or dialectic (are these different?), disagreement or problematicity? Is argumentation consensus-building or dissensus maintaining or does it have some other basic function? Or is essentialism with respect to argument a mistake? Is it better to think in terms of better and worse models of argument (and for what purposes)? Does the word 'argument' even as used by argumentation theorists have a single meaning or several senses? Is it worthwhile distinguishing between argument and argumentation?

(2) Types of argument

What different typologies of argument are there? For instance, argument as process or activity can be distinguished from argument as product or object; and argument as proof can be distinguished from argument as persuasion, and each of these types seem to contain sub-types What other such types are there? How are the apparently different types of argument related? Do all reduce to one or to a small number of basic types?

(3) The elements of arguments

Are arguments constituted by propositions (or sentences) or by speech acts (assertions, etc.), or something else, or does that answer depend on the perspective of the analysis? If so, are some perspectives superior or preferable to others?

(4) The structure of arguments

When interpreting arguments, what is the best way to conceive of their structure? What are the possible structural units: premises and conclusions? data, warrants and backing (Toulmin)? lines of argument? argument schemes? pro and con arguments? all of these? others? How may they fit together? See the possibilities in [Snoeck Henkemans, 1992]; but are there others still?

(5) Argumentation schemes (also known as schema and schemata)

The recent literature on argument schemes [Kienpointer, 1992; Snoeck Henkemans, 1992; Freeman, 1991; Walton, 1996b] raises questions about the nature of argument schemes, their types and their classification. Indeed, what exactly is an argument scheme (as distinct from an argument form, such as modus ponens)? Are schemes purely descriptive or purely normative, or can they be both? How is to be decided that an argument scheme has been characterized correctly? What is the role or function of argument schemes? Indeed, what motivates the theory of argument schemes?

(6) Argumentation, argument, reasoning, inference, implication

How are these concepts properly to be understood, both in general, and in relation to one another? Johnson [1996] calls this The Network Problem.. How are inferences differentiated from implications? Can arguments be reduced to implications?

(7) Criteria of good argument: what are they and how are they derived/justified?

Theorists have distinguished various types of criteria that might be applied in the evaluation of arguments, among them alethic, dialectical and rhetorical criteria. How is each to be characterized in detail? Are these truly distinct or is reduction among them possible? If they are distinct, are they compatible? If not, is that because it is possible to take incompatible perspectives on any argument, or is it because at least one is mistaken or only one is correct? Are there other types of criteria besides these?

Related to the general issues noted above are issues about the particulars of argument assessment. The dialectical criterion of premise-adequacy seems to be acceptance, or acceptability. Is it enough that premises are acceptable to the interlocutors? If not, what other criteria make them worthy of acceptance? From the epistemic perspective, should premises be true, or is some weaker criterion, such as reasonable to believe or probable? Where does plausibility fit as a criterion? Considering rhetorical criteria for premises, what is the appropriate role of such standards as the ethos of the arguer, among others?

Again, the choice of one or another general perspective — epistemic, dialectical, rhetorical or some other — will play a role in identifying the preferred criterion or criteria to be used in assessing the strength of the support for the conclusion that is supplied by the premises. If relevance and sufficiency are used as criteria, an analysis of each is needed. For instance, is relevance an "on/off" concept or a matter of degree? Is relevance a primitive concept, or can it be analysed in terms of other components? If Walton's distinction between "local" relevance (the relevance of premises to their conclusion) and "global" relevance (the relevance of an argumentative response to the question at issue) is tenable, how is each correctly to be understood? How are the elements of a parallel distinction between local and global sufficiency to be unpacked? Is Johnson [1996] right that arguers violate a sufficiency obligation if they fail to respond to known or reasonably-likely objections to their claims? Do the concepts of relevance and sufficiency obtain a grip only when applied in domain-specific senses and contexts?

The issue of connection-adequacy raises the question of what types of premise-conclusion relationship are possible and appropriate, since presumably each by definition has its own distinctive adequacy criteria. Here is where the distinctions between logical implication or entailment, inductive support and conductive support come up for scrutiny. Are Rescher's [1977] "provisoed assertion", Scriven's [1987] "probative inference", and Walton's [1996c] "presumptive inference" different from each other and is one or all of these to be added to the list of types of support relationship?

(8) The role of fallacy in argument analysis

We have already said enough about controversies surrounding fallacy theory and the accounts of particular fallacies to document the plethora of issues in this theoretical neighbourhood.

(9) How to compare rival theories

Given the large number of issues, and the inevitable differences between theoretical perspectives, the question naturally arises whether one theory about any one of these issues is preferable to the others, which in turn raises the meta-theoretical question, how are such questions to be settled? See Johnson [1995].

(10) Argumentation in the disciplines

Are the arguments in physics different from those in anthropology? Is the argumentation found in literary criticism different from that which characterizes disputes among economists? In short, setting aside content, are there salient differences in the arguments and argumentation

to be found in different disciplines? Are the norms of argument different? If so, what are those differences and what is their significance for the understanding of argumentation and for the analysis and assessment of arguments?

(11) Argumentation in the forum/public arena

Is the argumentation of the polis — the argumentation about political and social issues, about public policy, to be found in newspapers and magazines, on television and radio, on the hustings and in legislatures — of a distinct sort? What (if any) are the noteworthy features of such argumentation, both as it exists in this or that society, and also as it might exist in some more or less ideal version? Are there norms which can be identified as applying to this sort of argumentation, and if so, are these internal to the activity of argumentation in the public arena or do they inevitably presuppose some political or social philosophy?

(12) The relation between formal and informal logic

We have said that formal and informal logic are not incompatible. That comment is minimally informative about their relationship. Do they simply have different subject-matters? Can either serve the other in some capacity or another? How are they related? (See [Johnson, 1996; Massey, 1998].)

These, then, are some of the questions that we see as presently either not answered at all, or (more commonly) not answered to everyone's satisfaction. There are no doubt others that have escaped our attention.

VI. Resources

1 INTRODUCTION

A scholarly field consists of a community of scholars, communication among its members, and a disseminated literature. As informal logic has developed as a scholarly field — in distinction from a subject-matter for introductory logic, reasoning and critical thinking courses — it has developed such a scholarly infrastructure. Its components include journals, anthologies and monographs, conferences, and educational programs. The emergence of this thriving community in a bare quarter-century is impressive. What has not yet been developed is a market in academic departments for specialists in informal logic. Hence those who take up issues in informal logic as research questions come to the field with their primary or specialized training in other areas of philosophy, and those who design their entire research programs around informal logic are few, and tend to be advanced in their academic careers.

2 JOURNALS, ANTHOLOGIES AND MONOGRAPHS

Over the past twenty and more years, there has been a proliferation of literature on informal logic. Some who are ignorant of the field write as if the latest significant work is Hamblin's *Fallacies* [1970]. However serious scholars know that the literature has expanded and continues to proliferate to such an extent that it is becoming impossible for any one person to keep abreast of all the research being published.

The journal *Informal Logic* has been the journal of record in the field since 1983, however articles on topics in or related to informal logic also appear regularly in other journals, notably: *Argumentation* (founded 1986), *Philosophy and Rhetoric, Argumentation and Advocacy* (the journal of the American Forensic Association), and *Inquiry: Critical Thinking Across the Disciplines* (founded in 1988). In addition, as Hansen's (1995) bibliography shows, articles on informal logic topics are to be found distributed among some three dozen other academic journals.

In addition, the anthology on fallacies edited by Hansen and Pinto [1995] and on the history of informal logic edited by Walton and Brinton [1997], the voluminous proceedings of the four ISSA conferences (see below), the proceedings of numerous other conferences (see below), the monographs by Siegel [1988], Freeman [1991], Gilbert [1997] and Walton (see References), the collections by Govier [1987], Johnson [2000] and Tindale [1999] plus the numerous monographs and anthologies in the cognate fields of argumentation studies also constitute a significant literature.

3 CONFERENCES

The communication that sustains an academic field thrives on the face-to-face conversations and exchanges that conferences provide so efficiently. There have been numerous conferences devoted exclusively to informal logic or with sessions on informal logic over the past two decades.

The initial conferences on informal logic were held at the University of Windsor (Windsor, Ontario, Canada) in 1978, 1983 and 1989. We now mention, without any pretense at completeness, some of the other conferences that have been held in recent years.

Canada. McMaster University hosted a conference on teaching informal logic and critical thinking in universities in 1988, and a conference on relevance in argumentation in 1991. The Ontario Society for the Study of Argumentation held conferences at Brock University devoted largely to informal logic and argumentation theory in 1995, 1997 and 1999, with the Proceedings of the last two available on CD-ROM..

United States. Conferences with papers on both informal logic and critical thinking have been held at Sonoma State University (Rohnert Park,

California) yearly from May 1981 through August 1997; and at Christopher Newport College (Newport News, Virginia) each year from 1984 to 1988. Oakton Community College (DesPlaines, Illinois) hosted five conference on critical thinking from 1988 to 1992 at which sessions were devoted to informal logic. A conference on recent research on critical thinking and informal logic was held at George Mason University (Fairfax, Virginia) in 1995.

Europe. Although their organizing topic has been argumentation in general, the following conferences have welcomed many papers in informal logic. First and foremost, in The Netherlands, the quadrennial International Society for the Study of Argumentation conferences at the University of Amsterdam in 1986, 1990, 1994 and 1998 have seen scores of informal logic papers presented (see their proceedings, edited by van Eemeren, Grootendorst, Blair and Willard). Also in the Netherlands, there were papers on informal logic presented at a conference on norms in argumentation at the University of Utrecht in 1998 and a conference on logic and politics at Groningen in 1990. In the United Kingdom, the first British Conference on Informal Logic and Critical Thinking was held at the University of East Anglia in 1988. In France, there were informal logic papers at the Colloque de Cerisy at Cerisy-la-Salle in 1987, and at a conference on Commonplaces, Topoi, Stereotypes and Clichés at the Université Lumière in Lyon in 1992.

Other. The Association for Informal Logic and Critical Thinking has, since its inception in 1983, organized sessions on informal logic in conjunction with the annual meetings of the Eastern, Central and Pacific divisions of the American Philosophical Association, and at some of the annual meetings of the Canadian Philosophical Association.

4 EDUCATIONAL PROGRAMS

A field of scholarship flourishes as long as there is a well-defined problematic and a critical mass of scholars sufficiently intrigued to work on it. Concerns have been expressed (in conversation) that the informal logic community needs to recruit actively to ensure that younger scholars succeed the present aging cohort of active scholars.

To our knowledge, there is no undergraduate program of study devoted to informal logic. Nowhere is it possible to make informal logic the major concentration of study for a bachelor's degree. Virtually every college and university in the United States and Canada offers an introductory-level course that includes some informal logic, but we know of only a handful of upper-level undergraduate courses. There is another handful of courses at the M.A. or Ph.D. levels, mostly in Canada (at McMaster University, the University of Windsor, the Univeristy of Lethbridge and York University). It is possible to write a Master's or Ph.D. dissertation on a topic in informal logic at a very few universities.

5 IMPLIED CHALLENGES

In our view, the field of informal logic faces two major, and related, practical challenges. First, it needs to find the support and resources to mount graduate level instruction, so that more young scholars become aware of its problematic and its literature, and so that they make informal logic one area of concentration in their research programs. Second, it needs better to penetrate the broad scholarly establishment, so that its theoretical findings become known and better reflected both in the general understanding of arguments in philosophy and also in undergraduate instruction.

VII. Implications Beyond Informal Logic

1 INTRODUCTION

As a new normative theory of argument, informal logic has implications well beyond the "critical thinking" course classroom. If its insights are correct, then there should be a significant shift within philosophy. It would spell the end of the current privileging of deductive implication and formal methods — not their replacement, but their removal from their present pedestal — and a concomitant change in how philosophers understand reasoning and argument. Those changes should have a ripple effect on other disciplines that currently rely on philosophy's authority for their models of logic and reasoning. And insofar as reasoned argument plays a role in the world at large, informal logic's enriched understanding of its nature and norms holds the promise of importance beyond philosophy and the academy.

2 IMPLICATIONS FOR PHILOSOPHY

2.1 The end of deductivism

Philosophical reasoning and argument need informal logical analysis. It is a contentious philosophical position, not a simple fact, that philosophical reasoning and argumentation ought to employ only inferences intended to be deductively valid. We take Ryle to be describing a legitimate state of affairs, not lamenting a failing, when he wrote:

> Whether a given philosophical argument is valid or fallacious is, in general, itself a debatable question. Simple inspection cannot decide. More often it is a question of whether the argument has much, little or no force. [Ryle, 1954, p. 112]

Perhaps the most important contribution of informal logic is that it helps to complete the reorientation begun by the pragmatists who took issue with

the classical (Platonic/Cartesian) theory of knowledge. Their work can be seen as an attempt to reconceptualize knowledge according to the model of the empirical sciences. Work in informal logic can be seen as an somewhat parallel attempt to reconceptualize the relation of argument to logic. So Perelman and Olbrechts-Tyteca [1958, 14] see argumentation as requiring attention to the psychological and social conditions of discourse, in distinction from the decontextualized axiomatic systems of logic, and Toulmin [1958] sees argument as better modelled on legal interactive discourse than on abstract geometrical or mathematical systems. The informal logic critique rejects "deductivism" as a general theory of argument. It is still possible, to be sure, for the discipline of philosophy to insist that it is a special case, like mathematics, and that all good philosophical arguments must be deductively valid. Whether they have been, historically, is an empirical question that has to our knowledge never been studied. But even if they have not, nothing in principle prevents the imposition of deductivism as the norm for philosophical argument. That norm would be supported by the doctrine that the only genuinely philosophical truths are necessary truths. However, if informal logical has successfully questioned the *presumption* that no deductively invalid argument can be sound, then it has at least opened the door to the possibility that there can be good, invalid philosophical arguments, paradoxical though that sounds. Or, to put the point even more cautiously, accepting the informal logic critique raises the question whether some good philosophical arguments might be technically "unsound". We might then look and see whether all proper philosophical arguments must be proofs, or that for any argument in philosophy to be meritorious, it must be conclusive.

In our own opinion, progress in philosophy is highly dialectical, with arguments published after careful development, only to be found wanting by ingenious counterexamples, or at least to be opposed by cogent-seeming objections, which are themselves put into question by yet other objections to them. Thus philosophical "debates" are in fact extended dialectical exchanges, with few arguments found to be decisive or conclusive — and certainly not the philosophically more interesting ones. The *cogito* is a brilliant argument, but what does it prove? Discussions of it have filled volumes and occupied careers. Our suggestion is that philosophy can learn about its own argumentative methodology by attention to many of the issues that constitute the problematic of informal logic.

2.2 Logic and of the theory of reasoning

By identifying reasoning with inferences whose premises deductively imply their conclusions, and then taking the view that logic is essentially deductive logic, it was possible to believe that logic is in fact identical with the theory of reasoning. But once we adopt a broader view of both logic and reasoning,

it becomes clear that there is more involved in the construction of a theory of reasoning than formal deductive logic can provide [Finocchiaro, 1984; Johnson, 1996]. If we are right, philosophical education needs to change its standard story about argument, reasoning, and logic.

Another implication of the work in informal logic has been to make it clear that logic is still developing. Informal logic is not the only indicator of this fact. Other developments, such as fuzzy logic developed by Zadeh [1975] and dynamic logic developed by van Benthem [?], also attest to it. Yet, as mentioned, the vast body of philosophical education in remains untouched by the findings of informal logic. As a consequence, serious mis-education continues.

2.3 The re-evaluation of formalism

Another result of informal logic has been to challenge the strong attachment to formalism and all that goes with it: algorithms, proof procedures, model theories, and so on. It is not then just the deductive bias inherent in traditional logic that informal logic has helped reveal; it is also the bias in favour of certain kinds of formalism. This is what Toulmin is referring to when he says:

> From the mid 17th century, Modern Philosophers regarded the formal issues a central — not least, because they would be discussed in general, "decontextual" terms. So, logic became equated the formal logic. [1992, 4]

Here it is crucial to remember, as pointed out above, that there are different senses of the term, 'formal.' [Barth and Krabbe, 1982; Johnson and Blair, 1991]. This realization opens the door to seeing that "informal" logic is not a contradiction or oxymoron, as some [Hintikka, 1989] have alleged. Informal logic is in no way opposed to rules and procedures, to criteria and their application, or, where appropriate, to sets of logically necessary and sufficient conditions. It is a question of which rules, procedures and criteria, and when such conditions are needed. Here informal logic is informal because it rejects the logicist view that logical form (à la Russell) holds the key to understanding the structure of all arguments; and also the view that validity is the only appropriate criterion of logically good argument.

Another way of looking at this point is by seeing informal logic as allied with the movement to make logic more empirical, and less a prioristic [Toulmin, 1958; Barth, 1985; Finocchiaro, 1989; Weinstein, 1990]).

3 IMPLICATIONS FOR OTHER DISCIPLINES, PARTICULARLY COGNITIVE STUDIES AND COMPUTER SCIENCES

In a great many disciplines, formal deductive logic remains the model of logic and of argument. Time and time again one sees researchers hamstrung by the assumption that arguments are all deductive and that validity is the appropriate criterion for evaluating them. Yet the degree that one's interest turns to practical reasoning, one discovers ever so many ways in which exclusive reliance on the model handed down by formal deductive logic is inadequate.

For one thing, the formal concept of validity does not capture the kind of relationship we expect between premises and conclusion. The following is cited as a valid argument:

(S1) If the moon is made of green cheese, then arithmetic is complete.
(but) The moon is not made of green cheese.
(so) Arithmetic is complete.

To many, this example seems specious and certainly far removed from anything one would expect to find in the realm of practical reasoning. The problem here involves the representation of "if ..., then ..." by the truth-functor known as "material implication".

The unhappiness of this result led some logicians to pursue a better operator. In this quest, C.I. Lewis developed the systems of strict implication, and these systems are responsible for the development of modal logic [Lewis and Langford, 1935]. But strict implication as a rendering of the conditional which lies at the heart of logic and argument has its shortcomings. Later, Belnap and Anderson attempted to develop relevance logic [Anderson and Belnap, 1975]. By all accounts, it, too, fails to capture formally the required sense of "if ..., then ..." Thus, the attempt to render formally the entailment relationship (the soul of formal deductive logic's idea of an argument) turns out to be a *will o' the wisp*.

The shortcomings of the traditional view are slowly becoming better known within the artificial intelligence (AI) community. Thus Gabbay [1995] writes:

> The traditional modelling of an argument, as assumptions leading to a conclusion, is too narrow. There is a need to see the argument within a changing and evolving context, involving not only deduction but also other mechanisms such as abduction, action updates and consistency maintenance. (309)

Yet many in the AI community, apparently unaware of this shift, remain under the influence of the traditional view. For instance, Hendricks, Elvang-Gorannson and Pedersen [1995] write:

> Arguments are constructed using natural deduction over a propositional database. Arguments are pairs (δ, p) where δ is the set of facts in the database from which the conclusion of the argument, p is derived. (361)

Problems multiply when the realization is added that many arguments are not deductive in character. One might broaden the notion of induction to cover the rest of the territory, but the result is an unpalatable dilemma. Either *prima facie* strikingly different types of reasoning and argument have somehow to be assimilated in order to fit a single concept of induction, or else the concept of induction gets stretched so far out of shape as to be rendered useless to denote any one particular interesting pattern of reasoning or argument. Consider the following cases in point:

(S2) 99 (and) That bird is a raven.
(so) Probably that bird is black,

(S3) Tweety is a bird.
(and) Birds can fly.
(so) Probably Tweety can fly.

(S4) Candidate Able is well ahead in a recent poll.
(and) The pollster is reputable.
(and) The vote is tomorrow.
(so) Probably candidate Able will win.

(S5) You promised to take your son to a movie this afternoon.
(so) Presumably, you ought to do so.

It seems to us clear that each of these examples has quite distinct properties, and certainly the difference between (S2) and (S5) is striking. The prospect of making the case that they are all versions of a single type of reasoning or argument does not seem promising. Given their distinctiveness, nothing useful or illuminating is conveyed by gathering them all under the label "inductive".

Another possibility is to look for some third type of premise-conclusion link. This quest has led some, Hitchcock [1983] and Govier [1987], to name but two, to embrace (following Wellman [1971]) what has been called "conductive inference". This is reasoning or argumentation in which the premises furnish "good reasons" for the conclusion. More recently, Walton [1996c] used a concept he calls "presumptive" reasoning and arguments, in which the premises create a presumption in favour of the conclusion.

These examples of reasoning and argument that do not appear, on the face of it, to be deductively valid, and those that seem, on the face of it, not to fit standard inductive models or patterns either, raise the possibility that there may be another type of premise-conclusion relationship (or other

types), besides those modelled by entailment and by inductive strength. They thus challenge those who come to logic from other disciplines to not limit their examination of the field to formal deductive logic and the logic of probability.

4 IMPLICATIONS FOR THE LIFE-WORLD

Finally, it can be argued that the practice of argumentation has fallen on hard times in this culture. Serious observers have noted a decrease in literacy skills in the culture, the level of public debate seems at an all-time low, and public rhetoric is dominated by the confessional mode of television "talk-shows" (see Postman's *Amusing Ourselves to Death* 1985, for a pessimistic analysis). Where in all this is the practice of argumentation to be cherished and nurtured, if not in the Academy? The teaching of high standards of argument interpretation, evaluation and critique that have practical application is the goal of the pedagogical side of informal logic.

Outside the Academy, in what some would call the "life-world", recent years have witnessed the withering away of the old world order. The post Second World War coalitions existing under the threat of military force and power are now everywhere giving way to new alignments. But these are not always based on common interest and rational persuasion. The Balkan states, the middle East, many parts of Africa, and the India-Pakistan subcontinent, are noteworthy examples. The prospects for peace and prosperity in the human community await the understanding that the only force we can tolerate is "the force of the better argument". Yet, paradoxically, it seems that just when there has never been a greater need for argumentation in the life world, never has it been in greater danger as a cultural practice in the very societies premised on its healthy operation, the democracies. Thus, now more than ever, we the academic community, and particularly those committed to the study and healthy practice of everyday argumentation, have something to contribute in educating the world.

It is for these reasons that, in its commitment to the development of better theories of argumentation, informal logic has an important service to render not merely to the theory of reasoning and to the academy, but also to the life-world.

VIII. Summary

Since we believe that a scholarly field consists of its problematic and its history, which are inextricably intertwined, we have provided in this chapter a sketch of the problems constituting the subject matter of informal logic, placed in their historical context. Informal logic as a field may be traced most directly to Hamblin's *Fallacies* [1970] and Kahane's *Logic and*

Contemporary Rhetoric [1971], although inspiration was later gained from the earlier work of Toulmin [1958], Perelman and Olbrechts-Tyteca [1958] and others. Hamblin and Kahane symbolize the two strands of the field, theoretical and pedagogical, the combination of which led initially to its approach to logic and argument, its discomfort with various assumptions of the time, and the directions its theorizing took to start with.

We have proposed that informal logic as a field emerged from a dissatisfaction with formal deductive logic as the instrument of choice for instruction in the analysis and evaluation of arguments and as the theory of choice to account completely for the norms of argument and argumentation. Simultaneously, various social and demographic forces created a benign context in anglophone North American colleges and universities for informal logic's development. We traced the emergence from the logic texts at the end of the third quarter of the century of a shift to a focus on the analysis and evaluation of arguments in use and in context. We identified informal logic theory as the enterprise of supplying the norms and their theoretical framework appropriate to these tasks. And we set out in some detail the various approaches that have been taken to date to stock those particular theoretical shelves.

We ended the paper with a list of the issues that are currently open and receiving attention in the informal logic literature, with an account — for those new to the field — of the scope of the activity in the field, and with some speculations on the significance of informal logic beyond its own borders.

ACKNOWLEDGEMENTS

In writing this article we have drawn on work that has appeared in print or been presented in other places: Chapter 1, "Informal Logic, Past and Present", in *New Essays in Informal Logic* [Johnson and Blair, 1994a]; Chapter 6, Informal Logic and Critical Thinking, in *Fundamentals of Argumentation Theory* [Eemeren *et al.*, 1996]; Chapter 12, "Informal Logic in the Twentieth Century", in *Historical Foundations of Informal Logic* [Walton and Brinton, 1997]; and "Informal Logic: An Overview", a paper presented at the 20th World Congress of Philosophy, Boston, August 10–16, 1998.

BIBLIOGRAPHY

[Allen, 1998] D. Allen. Should we assess the basic premises of an argument for truth or acceptability? In: Hansen, Tindale and Colman (Eds). *Argumentation & Rhetoric*, St. Catharines, Ontario: OSSA. (CD-Rom), 1998.

[Anderson and Belnap, 1975] A. R. Anderson and N. D. Belnap, Jr. *Entailment: The Logic of Relevance and Necessity.* Princeton: Princeton University Press, 1975.

[Bar-Hillel, 1969] Y. Bar-Hillel. Formal logic and natural languages. In *Foundations of Language V*, 15ff, 1969.

[Barth, 1982] E. M. Barth. A normative-pragmatical foundation of the rules of some systems of formal$_3$ dialects. In [Barth and Martens, 1982, pp. 159–170], 1982.

[Barth, 1985] E. M. Barth. A new field: Empirical logic/bioprograms, logemes and logics as institution. *Synthese*, 63: 376–388, 1985.

[Barth and Krabbe, 1978] E. M. Barth and E. C. W. Krabbe. Formele$_3$ dialectiek: Instrumenten ter beslechting van conflicten over geuite meningen. *Spektator, Tdschrift voor Neerlandistiek* 7: 307–341, 1978.

[Barth and Krabbe, 1982] E. M. Barth and E. C. W. Krabbe. *From Axiom to Dialogue*. Berlin/New York: De Gruyter, 1982.

[Barth and Krabbe, 1992] E. M. Barth and E. C. W. Krabbe, eds. *Logic and Political Culture*. North Holland: Royal Netherlands Academy of Arts and Sciences, 1992.

[Barth and Martens, 1982] E. M. Barth and J. L. Martens, eds. *Argumentation: Approaches to Theory Formation*. Amsterdam: John Benjamins B.V, 1982.

[Barry, 1992] V. E. Barry. *The Critical Edge: Critical Thinking for Reading and Writing*. Fort Worth, TX: Harcourt Brace Jovanovich College Publishers, 1992.

[Beardsley, 1955] M. Beardsley. *Practical Logic*. Englewood Cliffs, NJ: Prentice-Hall, Inc, 1955.

[Beardsley, 1966] M. Beardsley. *Thinking Straight*. 4th edition. Englewood Cliffs, NJ: Prentice-Hall, Inc, 1966.

[Benthem, 1994] J. van Benthem. *Dynamic Logic and the Flow of Information*. Amsterdam: Institute for Language, Logic and Computation, 1994.

[Biro, 1992] J. Biro and H. Siegel. Normativity, argumentation and an epistemic theory of fallacies. In: Frans H. van Eemeren *et al.*, eds. Vol. 1: *Argumentation Illuminated*, pp. 85–103, 1992.

[Black, 1946] M. Black. *Critical Thinking*. 2nd edition, 1952. Englewood Cliffs, NJ: Prentice-Hall, 1946.

[Blair, 1992] J. A. Blair. What is the right amount of support for a conclusion? In van Eemeren *et al.*, 1992b.

[Blair, 1995] J. A. Blair. The place of teaching informal fallacies in teaching reasoning skills or critical thinking. In [Hansen and Pinto, 1995, pp. 328–338], 1995. 1995, 328-338.

[Blair and Johnson, 1980] J. A. Blair and R. H. Johnson, eds. *Informal Logic*, The First International Symposium. Inverness, CA: Edgepress, 1980.

[Blair and Johnson, 1987] J. A. Blair and R. H. Johnson. Argumentation as dialectical. *Argumentation* 1: 41–56, 1987.

[Brinton, 1985] A. Brinton. A rhetorical view of the ad hominem. *Australasian Journal of Philosophy*, 63: 50–63, 1985.

[Carney and Scheer, 1964] J. D. Carney and R. K. Scheer. *Fundamentals of Logic*.(3rd edition, 1980. New York: Macmillan Publishing Co., Inc. and London: Collier Macmillan Publishers, 1964.

[Cohen, 1989] L. J. Cohen. An Introduction to the Philosophy of Induction and Probability. Oxford: Clarendon Press, 1989.

[Cohen, 1995] M. Cohen. Probability. In Ted Honderich, ed., *The Oxford Companion to Philosophy*, 720. Oxford: Oxford University Press, 1995.

[Conway, 1991] D. A. Conway. On the distinction between convergent and linked arguments. *Informal Logic*, 8, 145–58, 1991.

[Copi, 1954] I. M. Copi. *Introduction to Logic*. 10th edition, 1998, with Carl Cohen. New York: MacMillan, 1954.

[Copi, 1986] I. M. Copi. *Informal Logic*. New York: MacMillan, 1986.

[Damer, 1987] T. E. Damer. *Attacking Faulty Reasoning*. 2nd edition. Belmont, CA: Wadsworth, 1987.

[Dewey, 1910/1991] J. Dewey. *How We Think*. Orig. Publ. Lexington, MA: D.C. Heath; republished Buffalo, NY: Prometheus Books, 1991.

[Dewey, 1929] J. Dewey. *The Quest for Certainty*. New York: Minton, Balch, 1929.

[Eells, 1995] E. Eells. Probability. In Robert Audi, (Ed)., *The Cambridge Dictionary of Philosophy*, 649–651. Cambridge: Cambridge University Press, 1995.

[Eemeren and Grootendorst, 1983] F. H. van Eemeren and R. Grootendorst. *Speech Acts in Argumentative Discussions*. Dordrecht: Foris, 1983.

[Eemeren and Grootendorst, 1992] F. H. van Eemeren and R. Grootendorst. *Argumentation, Communication and Fallacies, A Pragma-Dialectical Perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates, 1992.

[Eemeren *et al.*, 1987] F. H. van Eemeren, R. Grootendorst, J. A. Blair, C. A. Willard, eds. Conference on Argumentation 1986, in three volumes. *Argumentation: Across the Lines of Discipline; Argumentation: Perspectives and Approaches; Argumentation: Analysis and Practices*. Dordrecht-Holland/Providence-U.S.A.: Foris Publications, 1987.

[Eemeren *et al.*, 1992a] F. H. van Eemeren, R. Grootendorst, J. A. Blair, C. A. Willard, eds. *Argumentation Illuminated*. (Selected papers from the Second International Conference on Argumentation.) Amsterdam: SicSat, 1992.

[Eemeren *et al.*, 1992b] F. H. van Eemeren, R. Grootendorst, J. A. Blair, C. A. Willard, eds. *Proceedings of the Second International Conference on Argumentation*. Amsterdam: SicSat. (Bound in two volumes, continuous pagination.), 1992.

[Eemeren *et al.*, 1995] F. H. van Eemeren, R. Grootendorst, J. A. Blair, C. A. Willard, eds. *Proceedings of the Third ISSA Conference on Argumentation. Vol. I, Perspectives and Approaches; Vol. II, Analysis and Evaluation; Vol. III, Reconstruction and Application; Vol. IV, Special Fields and Cases*. Amsterdam: SicSat, 1995.

[Eemeren *et al.*, 1999] F. H. van Eemeren, R. Grootendorst, J. A. Blair, C. A. Willard, eds. *Proceedings of the Fourth International Conference of the International Society for the Study of Argumentation*. Amsterdam: SicSat, 1999.

[Eemeren *et al.*, 1996] F. H. van Eemeren, R. Grootendorst and F. Snoeck Henkemans, eds. *Fundamentals of Argumentation Theory*. Mahwah, NJ: Lawrence Erlbaum Associates, 1996.

[Ennis, 1982] R. H. Ennis. Identifying implicit assumptions. *Synthese* 51: 61–86, 1982.

[Finocchiaro, 1980] M. Finocchiaro. *Galileo and the Art of Reasoning: Rhetorical Foundations of Logic and Scientific Method*. Dordrecht: D. Reidel Publishing Company, 1980.

[Finocchiaro, 1981] M. Finocchiaro. . Fallacies and the evaluation of reasoning. *American Philosophical Quarterly*, 18: 13–22, 1981.

[Finocchiaro, 1984] M. Finocchiaro. Informal logic and the theory of reasoning. *Informal Logic*, 6: 3–8, 1984.

[Finocchiaro, 1989] M. Finocchiaro. Methodological problems in empirical logic. *Communication and Cognition*, 22: 313–333, 1989.

[Fisher and Scriven, 1997] A. Fisher and M. Scriven. *Critical Thinking: Its Definition and Assessment*. Point Reyes, CA: Edgepress, 1997.

[Fogelin, 1978] R. J. Fogelin. *Understanding Arguments, An Introduction to Informal Logic*. 4th edition, 1991, with Walter Sinnott-Armstrong. New York: Harcourt Brace Jovanovich, 1978.

[Freeman, 1988] J. B. Freeman. *Thinking Logically, Basic Concepts for Reasoning*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[Freeman, 1991] J. B. Freeman. *Dialectics and the Microstructure of Argument: A Theory of Argument Structure*. Amsterdam: Mouton de Gruyter, 1991. (originally: Berlin/New York: Foris Publications, PDA 10.)

[Gabbay, 1995] D. M. Gabbay. Labelled deductive systems and the informal fallacies: a preliminary analysis. In Frans H. van Eemeren *et al.*, 1995.

[Georgacarakos and Smith, 1979] G. N. Georgacarakos and R. Smith. *Elementary Forma Logic*. New York: McGraw-Hill, 1979.

[Gilbert, 1997] M. Gilbert. *Coalescent Argumentation*. Mahwah, NJ: Lawrence Erlbaum Associates, 1997.

[Goldman, 1986] A. I. Goldman. *Epistemology and Cognition*. Cambridge: Harvard University Press, 1986.

[Govier, 1982] T. Govier. Who says there are no fallacies? *Informal Logic Newsletter*, 5: 2–10, 1982.

[Govier, 1985] T. Govier. *A Practical Study of Argument*. 4th edition, 1997. Belmont, CA: Wadsworth, 1985.

[Govier, 1987] T. Govier. *Problems in Argument Analysis and Evaluation.* Dordrecht-Holland/Providence-U.S.A.: Foris, 1987.

[Grice, 1975] P. Grice. Logic and conversation. In P. Cole and J. Morgan, (Eds.). *Syntax and Semantics*, Vol. 3 (New York: Academic Press.), 1975. Reprinted as Ch. 2 in Grice 1989a.

[Grice, 1989a] P. Grice. *Studies in the Way of Words.* Cambridge, MA: Harvard University Press, 1989.

[Grice, 1989b] P. Grice. Further notes on logic and conversation. Ch. 3 in 1989a. Grünbaum, Adolf and Salmon, Wesley C., (Eds.). *The Limitations of Deductivism.* Berkeley: The University of California Press, 1989.

[Hamblin, 1970] C. L. Hamblin. *Fallacies.* London: Methuen, 1970.

[Hansen, 1990] H. V. Hansen. An informal logic bibliography. *Informal Logic*, 12: 155–184, 1990.

[Hansen, 1995] H. V. Hansen. Selected bibliography of recent work on the fallacies. In Hansen and Pinto, 1995, 339–348.

[Hansen and Pinto, 1995] H. V. Hansen and R. C. Pinto, eds. *Fallacies, Classical and Contemporary Readings.* University Park, PA: The Pennsylvania State University Pressm 1995.

[Hansen and Tindale, 2000] booktitle = "Argumentation at the Century's Turn", publisher = "St. Catharines, ON: OSSA", note = "CD", editor = "Hans V. Hansen and Christopher Tindale", year = "2000",

[Harman, 1986] G. Harman. *Change in View, Principles of Reasoning.* Cambridge, MA: MIT Press (Bradford), 1986.

[Hendricks *et al.*, 1995] V. F. Hendricks, M. Elvang-Göransson and S. A. Pedersen. Systems of Argumentation. In Frans H. van Eemeren, et al., *Proceedings of the Third ISSA Conference, Volume III: Reconstruction and Applications*, 351–367, Amsterdam: SicSat, 1995.

[Hintikka, 1989] J. Hintikka. The role of logic in argument. *The Monist*, 72, 3–24, 1989.

[Hintikka and Bachman, 1991] J. Hintikka and J. Bachman. *What If ...? Toward Excellence in Reasoning.* Mountain View, CA: Mayfield Publishing Company, 1991.

[Hitchcock, 1983] D. Hitchcock. *Critical Thinking, A Guide to Evaluating Information.* Toronto: Methuen, 1983.

[Hitchcock, 1992] D. Hitchcock. Relevance, *Argumentation* 6, 251–70, 1992.

[Hitchcock, 1995] D. Hitchcock. Do the fallacies have a place in the teaching of reasoning skills or critical thinking? In Hansen and Pinto, 1995, 319–327.

[Johnson, 1981] R. H. Johnson. The new logic course: the state of the art in non-formal methods of argument analysis. Teaching Philosophy 4 (2): 123-143.

[Johnson, 1987] R. H. Johnson. The blaze of her splendours: Suggestions about revitalizing fallacy theory. *Argumentation*, 1, 239–253,m1987.

[Johnson, 1989] R. H. Johnson. Massey on fallacy and informal logic: a reply. *Synthese*, 80, 407–26, 1989.

[Johnson, 1990a] R. H. Johnson. Hamblin on the standard treatment. *Philosophy and Rhetoric*, 23, 153–67, 1990.

[Johnson, 1990b] R. H. Johnson. Acceptance is not enough: A critique of Hamblin. *Philosophy and Rhetoric*, 23, 271–287, 1990.

[Johnson, 1995] R. H. Johnson. Informal logic and pragma-dialectics: some differences. In Frans H. van Eemeren, et al., *Proceedings of the Third ISSA Conference, Volume I: Perspectives and Approaches*, Vol. I (pp. 3237–245), Amsterdam: SicSat, 1995.

[Johnson, 1996] R. H. Johnson. *The Rise of Informal Logic.* Newport News, VA: Vale Press, 1996.

[Johnson, 1999] R. H. Johnson. The problem of truth for theories of argument. Paper presented to the Fourth ISSA Conference, Amsterdam, June 1998. In Frans H. van Eemeren, et al., *Proceedings of the Fourth ISSA Conference*, Amsterdam: SicSat, 1999.

[Johnson, 2000] R. H. Johnson. *Manifest Rationality.* Mahwah, NJ: Lawrence Erlbaum Associates, 2000.

[Johnson and Blair, 1977] R. H. Johnson and J. A. Blair. *Logical Self-Defense.* 3rd edition, 1993. Toronto: McGraw-Hill Ryerson, 1977. (First United States edition: 1994. New York: McGraw-Hill.)

[Johnson and Blair, 1980] R. H. Johnson and J. A. Blair. The recent development of informal logic. In J. Anthony Blair and Ralph H. Johnson (Eds.). *Informal Logic, The First International Symposium*, 3–28. Inverness, CA: Edgepress, 1980.

[Johnson and Blair, 1985] R. H. Johnson and J. A. Blair. Informal logic: the past five years, 1978–83. *American Philosophical Quarterly*, 22(3), 181–196, 1985.

[Johnson and Blair, 1991] R. H. Johnson and J. A. Blair. Contexts of informal reasoning: Commentary. In James F. Voss, David N. Perkins and Judith W. Segal, (Eds.). *Informal Reasoning and Education*, 131–150. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc, 1991.

[Johnson and Blair, 1994a] R. H. Johnson and J. A. Blair. *New Essays in Informal Logic.* Windsor, ON: Informal Logic, 1994.

[Johnson and Blair, 1994b] R. H. Johnson and J. A. Blair. Informal logic, past and present. In Johnson and Blair, (Eds.). 1994a, 1–19, 1994.

[Johnson and Blair, 1996] R. H. Johnson and J. A. Blair. Informal logic and critical thinking. In van Eemeren, et al., *Fundamentals of Argumentation Theory*, pp. 163–188, 1996.

[Johnson and Blair, 1997] R. H. Johnson and J. A. Blair. Informal logic in the twentieth century. In Douglas Walton and Alan Brinton (Eds.). Historical Foundations of Informal Logic, pp.158–177, 1997.

[Joseph, 1906] H. W. B. Joseph. *An Introduction to Logic.* Oxford: The Clarendon Press, 1906.

[Kahane, 1969] H. Kahane. *Logic and Philosophy. 5th edition*, 1986. Belmont, CA: Wadsworth, 1969.

[Kahane, 1971] H. Kahane. *Logic and Contemporary Rhetoric.* 8th edition, 1997, with Nancy Cavender. Belmont, CA: Wadsworth, 1971.

[Kienpointer, 1992] M. Kienpointer. *Alltagslogik, Struktur und Funktion von Argumentationsmustern, Stuttgart-Bad Cannstatt: frommann-holzboog*, 1992.

[Kneale and Kneale, 1962] W. Kneale and M. Kneale. *The Development of Logic.* Oxford: Clarendon Press, 1962.

[Lehman *et al.*, 1988] D. R. Lehman, R. O. Lempert and R. E. Nisbett. The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist*, 43: 431–442, 1988.

[Lempereur, 1991] A. Lempereur, ed. *L'Argumentation, Colloque de Cerisy.* Liège: Mardaga, 1991.

[Lewis and Langford, 1935] C. I. Lewis and C. H. Langford. *Symbolic Logic.* 2nd edition, 1959. Publisher of 2nd edition: New York: Dover Publications, Inc, 1935.

[Little *et al.*, 1989] J. F. Little, L. A. Groarke and C. W. Tindale. *Good Reasoning Matters! A Constructive Approach to Critical Thinking.* 2nd edition, 1997, with Linda Fisher. Toronto: Oxford University Press, 1989.

[Maier, 1989] R. Maier, ed. *Norms in Argumentation, Proceedings of the Conference on Norms.* Dordrecht: Foris, 1989.

[Massey, 1975] G. Massey. Are there any good arguments that bad arguments are bad? *Philosophy in Context*, 4: 61–77, 1975.

[Massey, 1981] G. Massey. The fallacy behind fallacies. *Midwest Studies in Philosophy*, 6: 489–500, 1981.

[Massey, 1998] G. Massey. Informal Logic: Amorphic or Morphophobic? Paper delivered at the AILACT Session, meeting of the Central Division of the American Philosophical Association, May 1998.

[McPeck, 1981] J. McPeck. *Critical Thinking and Education.* New York: St. Martin's Press,1981.

[Nisbett, 1993] R. E. Nisbett, ed. *Rules for Reasoning.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1993.

[Perelman and Olbrechts-Tyteca, 1958] C. Perelman and L. Olbrechts-Tyteca. *La Nouvelle Rhétorique: Traité de l'Argumentation*. Paris: Presses Universitaires de France, 1958. (*The New Rhetoric: A Treatise on Argumentation*, tr. by John Wilkinson and Purcell Weaver, Notre Dame and London: University of Notre Dame Press, 1969.)

[Plantin, 1993] C. Plantin, ed. *Lieux Communs, Topoï, Stéréotypes, Clichés*. Paris: Éditions Kimé, 1993.

[Popper, 1959] K. Popper. *The Logic of Scientific Discovery*. London: Hutchinson & Co, 1959. (Tr. of Logik der Forschung, 1935.)

[Postman, 1985] N. Postman. *Amusing Ourselves to Death*. New York: Viking, 1985.

[Quine, 1952] W. V. O. Quine. *Methods of Logic*. Rev. ed. 1959. New York: Random House, 1952.

[Quine and Ullian, 1970] W. V. O. Quine and J. S. Ullian. *The Web of Belief*. 2nd edition, 1978. New York: Random House, 1970.

[Rescher, 1977] N. Rescher. *Dialectics, A Controversy-Oriented Approach to the Theory of Knowledge*. Albany: State University of New York Press, 1977.

[Ryle, 1954] G. Ryle. *Dilemmas*. Cambridge: Cambridge University Press, 1954.

[Scriven, 1976] M. Scriven. *Reasoning*. New York: McGraw-Hill, 1976.

[Scriven, 1987] M. Scriven. Probative logic: review and preview. In van Eemeren *et al*, eds. *Argumentation Across the Lines of Discipline*, 7–32. Dordrecht: Foris, 11987.

[Seech, 1993] Z. Seech. *Open Minds and Everyday Reasoning*. Belmont, CA: Wadsworth, 1993.

[Siegel, 1988] H. Siegel. *Educating Reason*. London: Routledge, 1988.

[Skyrms, 1966] B. Skyrms. *Choice and Chance*. Belmont, CA: Dickenson Publishing Co, 1966.

[Snoeck Henkemans, 1992] A. F. Snoeck Henkemans. *Analysing Complex Argumentation*. Amsterdam: SicSat, 1992.

[Sperber and Wilson, 1986] D. Sperber and D. Wilson. *Relevance, Communication and Cognition*. Cambridge, MA: Harvard University Press, 1986.

[Stebbing, 1931] S. L. Stebbing. *A Modern Elementary Logic*. London: Methuen, 1931.

[Stein, 1996] E. Stein. *Without Good Reason*. Oxford: Clarendon, 1996.

[Thomas, 1973] S. N. Thomas. *Practical Reasoning in Natural Language*. 3rd edition, 1997. Englewood Cliffs, NJ: Prentice-Hall, Inc, 1973.

[Tindale, 1999] C. Tindale. *Acts of Arguing: A Rhetorical Model of Argument*. State University of New York Press, 1999.

[Toulmin, 1958] S. Toulmin. *The Uses of Argument*. Cambridge: Cambridge University Press, 1958.

[Toulmin, 1992] S. Toulmin. Logic, rhetoric & reason: Redressing the balance. In van Eemeren *et al.*, eds. *Argumentation Illuminated*, 3–11. Amsterdam: SicSat, 1992.

[Toulmin *et al.*, 1978] S. Toulmin, R. Rieke and A. Janik. *An Introduction to Reasoning*. 2nd edition, 1984. New York: Macmillan Publishing Co., Inc./London: Collier Macmillan Publishers, 1978.

[Vorobej, 1995a] M. Vorobej. Hybrid arguments. *Informal Logic*, 17.2: 289–296 1995.

[Vorobej, 1995b] M. Vorobej. Linked arguments and the validity requirement *Argumentation*, 9.2: 291—304, 1995.

[Walton, 1982] D. N. Walton. *Topical Relevance in Argumentation*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1982.

[Walton, 1984] D. N. Walton. *Logical Dialogue Games and Fallacies*. Lanham, MD: University Press of America, 1984.

[Walton, 1987] D. N. Walton. *Informal Fallacies, Towards a Theory of Argument Criticisms*. Amsterdam/Philadelphia: John Benjamins Publishing Company, 1987.

[Walton, 1989a] D. N. Walton. *Informal Logic, A Handbook for Critical Argumentation*. Cambridge: Cambridge University Press, 1989.

[Walton, 1989b] D. N. Walton. *Question-Begging Argumentation*, 1989.

[Walton, 1991] D. N. Walton. *Begging the Question, Circular Reasoning as a Tactic of Argumentation*. New York: Greenwood Press, 1991.

[Walton, 1992a] D. N. Walton. *The Place of Emotion in Argument*. University Park, PA: The Pennsylvania State University Press, 1992.

[Walton, 1992b] D. N. Walton. *Plausible Argument in Everyday Conversation*. Albany, NY: New York State University Press, 1992.

[Walton, 1992c] D. N. Walton. *Slippery Slope Arguments*. Oxford: Oxford University Press, 1992.

[Walton, 1995] D. N. Walton. *A Pragmatic Theory of Fallacy*. Tuscaloosa and London: University of Alabama Press, 1995.

[Walton, 1996a] D. N. Walton. *Argument Schemes for Presumptive Reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates, 1996.

[Walton, 1996b] D. N. Walton. *Arguments from Ignorance*. University Park, PA: The Pennsylvania State University Press, 1996.

[Walton, 1996c] D. N. Walton. *Argument Structure, A Pragmatic Theory*. Toronto: University of Toronto Press, 1996.

[Walton, 1997] D. N. Walton. *Appeal to Pity*. Albany, NY: State University of New York Press, 1997.

[Walton, 1998a] D. N. Walton. *One-Sided Arguments: A Dialectical Analysis of Bias*. Albany: State University of New York Press, 1998.

[Walton, 1998b] D. N. Walton. *Ad Hominem Arguments*. Tuscaloosa and London: The University of Alabama Press, 1998.

[Walton, 1998c] D. N. Walton. *The New Dialectic: Conversational Contexts of Argument*. Toronto: University of Toronto Press, 1998.

[Walton and Brinton, 1997] D. N. Walton and A. Brinton, eds. *Historical Foundations of Informal Logic*. Aldershot: Ashgate, 1997.

[Walton and Krabbe, 1995] D. N. Walton and E. C. W. Krabbe. *Commitment in Dialogue, Basic Concepts in Interpersonal Reasoning*. Albany, NY: State University of New York Press, 1995.

[Wason and Johnson-Laird, 1972] P. C. Wason and P. N. Johnson-Laird. *Psychology of Reasoning: Structure and Content*. Cambridge, MA: Harvard University Press, 1972.

[Weinstein, 1990] M. Weinstein. Towards a research agenda for informal logic and critical thinking. *Informal Logic*, 12 (3): 121–143, 1990.

[Weinstein, 1994] M. Weinstein. Informal logic and applied epistemology. In Johnson and Blair (Eds.). *New Essays in Informal Logic*, 140–161. Windsor, Ontario: Informal Logic, 1994.

[Wellman, 1971] C. Wellman. Challenge and Response, Justification in Ethics. Carbondale and Edwardsville: Southern Illinois University Press/London and Amsterdam: Feffer & Simons, Inc, 1971.

[Wittgenstein, 1953] L. Wittgenstein. *Philosophical Investigations*. New York, The MacMillan Company, 1953.

[Woods, 1992] J. Woods. Who cares about the fallacies? In van Eemeren *et al.* eds. *Argumentation Illuminated*, 23–48, 1992.

[Woods, 1994] J. Woods. Sunny prospects for relevance? In Johnson and Blair, eds. *New Essays in Informal Logic*, 82–92, 1994.

[Woods, 1995] J. Woods. Fearful symmetry. In Hansen and Pinto, eds. *Fallacies*, 181–193, 1995.

[Woods and Walton, 1972] J. Woods and D. N. Walton. On fallacies, *Journal of Critical Analysis*, 4: 103–11, 1972. (Also in Woods and Walton (1989), 1–10.)

[Woods and Walton, 1982] J. Woods and D. N. Walton. *Argument: The Logic of the Fallacies*. Toronto: McGraw-Hill Ryerson, 1982.

[Woods and Walton, 1989] J. Woods and D. N. Walton. *Fallacies, Selected Papers 1972–1982*. Dordrecht/Providence: Foris Publications, 1989.

[Wreen, 1987] M. J. Wreen. Yes, Virginia, there is a Santa Claus. *Informal Logic*, 9 (1): 31–39, 1987.

[Yanal, 1991] R. J. Yanal. Dependent and independent reasons. *Informal Logic*, 13 (3): 137–144, 1991.

[Zadeh, 1975] L. Zadeh. *Fuzzy Sets and Their Applications to Cognitive and Decision Processes*. New York: Academic Press, 1975.

JON WILLIAMSON

PROBABILITY LOGIC

Practical reasoning requires decision-making in the face of uncertainty. Xenelda has just left to go to work when she hears a burglar alarm. She doesn't know whether it is hers but remembers that she left a window slightly open. Should she be worried? Her house may not be being burgled, since the wind or a power cut may have set the burglar alarm off, and even if it isn't her alarm sounding she might conceivably be being burgled. Thus Xenelda can not be certain that her house is being burgled, and the decision that she takes must be based on her degree of certainty, together with the possible outcomes of that decision.

If Xenelda, or X for short, uses classical logic to make a decision, she will not get very far. Classical logic has no explicit mechanism for representing the degree of certainty of premises in an argument, nor the degree of certainty in a conclusion, given those premises. X must look for a logic that can represent uncertainty.

Here we will look to probability as a representation of uncertainty, and see how a logic of practical reasoning might involve probability, in order to help agents like X. The plan is this: first we shall recall the standard definition of probability, and the standard interpretations of this formal definition. Next we shall look at attempts to incorporate probability into a logic, and go on to investigate a practical question, namely inference.

1 PROBABILITY AND ITS INTERPRETATIONS

Probability has a rather technical mathematical characterisation. Given an arbitrary non-empty space Ω and a σ-field $\mathcal{F}$ of subsets of Ω (that is, a nonempty class of subsets of Ω closed under complement and countable unions), $p : \mathcal{F} \longrightarrow \mathbf{R}$ is a *probability measure* if (for all $a, b, a_1, a_2, \ldots \in \mathcal{F}$)

M1: $0 \leq p(a) \leq 1$;

M2: $p(\emptyset) = 0$ and $p(\Omega) = 1$;

M3: if $a_1, a_2, \ldots$ are disjoint then $p(\bigcup_{i=1}^{\infty} a_i) = \sum_{i=1}^{\infty} p(a_i)$.

A new function, *conditional probability* is induced by probability according to the following definition:

MC: if $p(a) > 0$ then $p(b|a) = p(a \cap b)/p(a)$.

For example, let B stand for burglary and B' for no burglary. Suppose $\Omega = \{B, B'\}$, and $\mathcal{F} = \{\emptyset, \{B\}, \{B'\}, \{B, B'\}\}$. Then the axioms require

that $p(\{B\}) + p(\{B'\}) = 1$ and the definition of conditional probability ensures that $p(\{B\}|\{B'\}) = 0$.

This gives a formal definition of probability,[1] but it doesn't tell us what probability means. Probability has a number of standard interpretations, and we shall take a brief look at these now.[2] As a starting point the elements of the domain of a probability measure are usually called events. Just what the events are depends somewhat on the interpretation of probability.

1.1 Subjective degrees of belief

Ramsey[3] and de Finetti[4] interpreted probability as degree of rational belief. According to this interpretation an agent, X for example, has a belief function p_X^τ at time τ over the domain of events, which are single-case (that is, unrepeatable). X's house being burgled that morning would qualify as such an event. Thus for $a \in \mathcal{F}, p_X^\tau(a)$ measures X's degree of belief at τ in the occurrence of event a. In our example, $p_X^\tau(\{B\})$ represents X's degree of belief at time τ that a burglary has taken place that morning. One can determine X's degrees of belief by analysing how she is prepared to bet:

- $p_X^\tau(a) = x$ if and only if at time τ, X is willing to bet $x\Delta\Theta$ on event a occurring, with return $\Delta\Theta$ if a does occur, where $\Delta\Theta$ is an unknown stake (either monetary or in terms of some measure of utility) which may depend on $p_X^\tau(a), \Theta \in \mathbf{R}_{\geq \mathbf{0}}$ being the magnitude and $\Delta = \pm 1$ the direction of the stake; and

- $p_X^\tau(b|a) = x$ iff at τ, she is prepared to bet $x\Delta\Theta$ on b occurring, with return $\Delta\Theta$ if both a and b occur but with the bet being called off if a fails to occur.

X's belief function p_X^τ is deemed to be *coherent* if no stake-maker can choose stakes which make X lose money whatever happens. By the Dutch book argument,[5] p_X^τ is coherent if and only if it is a probability measure. Thus probability measures are coherent belief functions. This gives a subjective interpretation of probability in the sense that probabilities are associated with a subject's state of knowledge or belief, as opposed to attaching directly to the physical world as stipulated by objective interpretations.

The notion of coherence can be used to argue for further rational constraints on X's belief function. For example one can apply a diachronic

---

[1][Kolmogorov, 1933] was a key pioneer behind the mathematical theory of probability — see [von Plato, 1994] for further historical details and [Billingsley, 1979] for a good exposition of the modern theory.

[2]See [Howson, 1995] for a more detailed survey.

[3][Ramsey, 1926].

[4][de Finetti, 1937].

[5][Ramsey, 1926] and [de Finetti, 1937]. See also [Williamson, 1999] regarding the axiom of countable additivity [*M3*].

Dutch book argument to show that if $p^{\tau}_X(a) > 0$ and between times τ and $\tau+1$ X comes to learn only of the occurrence of event a, then her belief function should change via *Bayesian conditionalisation*:

- $p^{\tau+1}_X(b) = p^{\tau}_X(b|a)$.

Thus one can bolster the subjective theory by adding further rationality requirements — although proponents of the subjective interpretation often disagree as to which extra principles should be adopted.[6]

Another point of disagreement boils down to attitudes toward other interpretations of probability. *Strict subjectivists* like de Finetti argue that subjectivism is the only viable interpretation of probability, while others are tolerant of, or argue in favour of, one or more of the following objective interpretations.

1.2 Objective frequencies

Von Mises was the first to work through the idea that probabilities are measures of frequency.[7] According to this interpretation an event a in $\mathcal{F}$ can be repeatably instantiated, *a burglary* rather than *X's burglary that morning*, and the *frequency* of a may be defined as its limiting relative frequency in a *collective*. This collective is a denumerable sequence of mutually exclusive and exhaustive elements of $\mathcal{F}$. In our example a collective may look like $(\{B'\}, \{B\}, \{B'\}, \{B\}, \{B'\}, \{B'\}, \{B'\}, \ldots)$ and may be obtained from examining X's house each time its alarm sounds to see if it was burgled. Von Mises invoked two empirical laws, the first of which claimed that for any naturally occurring collective, the relative frequency of an event a in the first n places of the collective tends to a limit, the frequency of a, as n increases. Thus in the above collective the relative frequency of $\{B\}$ progresses $0, \frac{1}{2}, \frac{1}{3}, \frac{1}{2}, \frac{2}{5}, \frac{1}{3}, \frac{2}{7}, \ldots$ and is assumed to converge to a fixed limit. The second empirical law claims that this frequency is constant over all subsequences selected by recursive place selections from the collective — so for example if we form a new collective by taking all the even places of the original collective, the limiting frequency in the new collective is the same as that in the original collective.[8] From these empirical laws von Mises' deduced that frequency obeys the axioms of *finitely additive* probability, that is, the above axioms with [*M3*] replaced by finite additivity, which is the special case in which finitely many attributes are considered.[9]'

Popper noted that there are intuitive difficulties concerned with the collective of outcomes of rolls of a biased die interspersed with one or two rolls

[6]Those who advocate Bayesian conditionalisation usually call themselves 'Bayesians'.

[7][von Mises, 1928].

[8]See [von Mises, 1964] for the details.

[9]Note however that von Mises later included countable additivity as an extra stipulation.

of a fair die. While it makes sense to say that the frequency of a 5 is $\frac{1}{4}$ for such a collective, intuitively the frequency changes according to which die is rolled.[10] Thus von Mises' theory is often amended to a *propensity* theory.[11] This states that it is reasonable only to consider collectives generated by a repeatable experiment, such as rolling a particular die, and that once this move is made, the frequency of an event is dependent only on the generating experiment. In other words the frequency is constant over all collectives generated by the same repeatable experiment. This might be given the status of an empirical law, or, as with Popper, such a concept of probability may be taken as a scientific primitive.

Another version of the frequency theory takes issue with the law that says the relative frequency in a collective converges to a fixed limit. One can argue that one should only assume that relative frequency converges *in probability* to the limit, which means that for a small proportion of collectives (a set of measure 0) there will be no convergence.[12] Another frequency theory, *actual frequency*, takes issue with the demand that collectives be infinite, which usually fails for outcomes of real experiments, and defines frequency to be the relative frequency in the actual (often finite) sequence of outcomes.[13]

1.3 Objective chances

We obtain quite a different objective notion of probability if we demand that events in the domain of a probability measure are single-case (as they were in the subjective theory) as opposed to the repeatable events of the frequency theory. We can merely hypothesise that the mathematical theory has such an objective physical interpretation, and find ways to test this assertion in order to confirm or refute it. Thus while many versions of the last two interpretations offer an explicit definition of probability in terms of observable beliefs and frequencies respectively, this theory implicitly defines objective single-case probabilities, or *chances*, and requires some form of prediction method to test the definition.[14] There are two standard ways of drawing predictions from chances. One can claim that chances give rise to certain frequencies as experiments are repeated, and then test to see

---

[10] [Popper, 1983] pp. 352–356.

[11] See for example [Popper, 1972]. The idea behind propensity theories, like that behind frequency theories, goes back a long way — see [Peirce, 1910] paragraph 664.

[12] The mathematical theory of probability only implies convergence according to the laws of large numbers, and consequently one can only account for the phenomenon of frequency *from within* the mathematical theory if one assumes the weaker notion of convergence in probability. See [Kolmogorov, 1933], [Neapolitan, 1992].

[13] See [Williamson, 1999b].

[14] Note that Popper's formulation of the propensity theory, while not a single-case theory, made use of this type of implicit definition of probability rather than von Mises' operationalist definition involving empirical laws.

whether those frequencies are obtained. Thus if the chance of X's house being burgled this morning, given that her alarm is sounding, is $\frac{1}{3}$, then one can claim that if we form a collective by checking for burglary each time X's alarm sounds, the frequency of burglary will be $\frac{1}{3}$ in this collective. Alternatively one can claim that the chance at time τ of an event is the degree to which an agent should believe at τ that it will occur, were she to have at τ all the relevant information pertaining to the occurrence of the event. We might then test this claim by seeing if the agent can be made to lose money by betting according to chances. Thus in our example X should believe that her house is being burgled, given that her alarm is sounding and other relevant information, to degree $\frac{1}{3}$.[15]

Having sketched the concept of probability and its interpretations, we shall move on to the relationship between probability and logic, which is the central theme of this chapter. In the next section we will look at attempts to integrate probability and logic. Later we shall investigate the important practical problems to do with probabilistic representation and inference.

2 PROBABILITY AS LOGIC

The starting point for most theories that integrate probability and logic is to attach probability to logical statements rather than events. The motivation here is that logic operates on statements and so it would be natural if a probabilistic logic for practical reasoning were to do the same. A second motivation is that when we look at the application of the mathematical theory of probability, we see that probabilities are usually posited of a random variable taking a certain value. Such expressions are more naturally thought of as statements of the form $X = x$ than events of the form $\{\omega \in \Omega : X(\omega) = x\}$.[16]

Unfortunately, confusion is often the upshot of the move from events to sentences. The problem is that the literature contains a plethora of new axiomatisations of probability on sentences, few of which bear a clear resemblance the mathematical formulation of probability.[17] One can however make the link between probability on sentences and the mathematical theory more perspicuous, as follows.

[15] See [Mellor, 1971] and [Lewis, 1980] for detailed defences of the chance approach.

[16] [Scott and Kraus, 1966], p. 219.

[17] Thus it requires significant effort to work out how the new notions of probability relate to the mathematical notion — see [Roeper and Leblanc, 1999] for a glimpse of what is required. One reason for the abundance of axiomatisations is that probability theory is often used to give a new semantics for logic, whereby a statement is logically true if given probability 1 by all probability functions, in which case axiomatisations of probability must be *autonomous*, in that they must not themselves involve logical notions. [Popper, 1934] appendix *iv and [Field, 1977] provide examples of such axiomatisations and this approach. However, our task is to investigate probability as an extension rather than a replacement of logic, so we need not enter into the intricacies of these axiomatisations.

Consider a propositional language $\mathcal{L}$ involving a countable set of propositional variables $\{c_1, c_2, \ldots\}$. Sentences $\mathcal{S}^\infty\mathcal{L}$ are formed by applying the usual connectives $\neg, \vee, \wedge, \rightarrow, \leftrightarrow$ and allowing denumerable disjunctions and conjunctions $\bigvee, \bigwedge$ (such sentences are known as *infinitary*).[18] The usual semantics (including logical truth, logical implication $\models$ and logical equivalence $\equiv$) is available if we consider the (uncountable) infinity of truth functions. Then $p : \mathcal{S}^\infty\mathcal{L} \longrightarrow \mathbf{R}$ is a *probability measure* if (for all $\theta, \phi, \theta_1, \theta_2, \ldots \in \mathcal{S}^\infty\mathcal{L}$)

L1: $0 \leq p(\theta) \leq 1$;

L2: $p(\theta \wedge \neg\theta) = 0$ and $p(\theta \vee \neg\theta) = 1$;

L3: if $\theta_1, \theta_2, \ldots$ are mutually exclusive[19] then $p(\bigvee_{i=1}^\infty \theta_i) = \sum_{i=1}^\infty p(\theta_i)$;

and *conditional probability* is defined by:

LC: if $p(\theta) > 0$ then $p(\phi|\theta) = p(\theta \wedge \phi)/p(\theta)$.

Note that one consequence of the axioms is that if two sentences are logically equivalent, $\theta \equiv \phi$, then they have the same probability.

The clear connection between the mathematical and the logical axioms of probability is due to the formal fact that the *Lindenbaum algebra* $\mathcal{S}^\infty\mathcal{L}/\equiv$ is a σ-algebra which is isomorphic to a σ-field $\mathcal{F}$ of subsets of the space Ω of truth functions. Thus each sentence θ corresponds to an element a of $\mathcal{F}$ and under this mapping the logical axioms on $\mathcal{S}^\infty\mathcal{L}$ are equivalent to the mathematical axioms on $\mathcal{F}$.

This works according to the following construction. The space Ω of truth functions is just the space of binary sequences. A truth function ω is of the form $(t_1(\omega), t_2(\omega), \ldots)$ where $t_i(\omega) \in \{T, F\}, i = 1, 2, \ldots$, signifies the truth value of c_i. A *cylinder of rank* n is of the form $a = \{\omega : (t_1(\omega), \ldots, t_n(\omega)) \in H\}$, where $H \subseteq \{T, F\}^n$. The set of cylinders of all ranks is a field,[20] and we shall consider the σ-field $\mathcal{F}$ generated by this field. Now to each sentence θ in $\mathcal{S}^\infty\mathcal{L}$ there corresponds an element of $\mathcal{F}$ which contains just the truth functions that satisfy θ. This can be shown inductively. If θ is propositional variable c_n, take H as $\{T, F\}^{n-1} \times \{T\}$ and then the corresponding cylinder of rank n contains just the truth functions that satisfy θ. If θ is $\neg\phi$ and $a \in \mathcal{F}$ contains the satisfiers of ϕ, then the complement of a, a', which is also in $\mathcal{F}$, will contain the satisfiers of θ. The union of the satisfiers of θ and ϕ will be the satisfiers of $\theta \vee \phi$, and a countable union is required for $\bigvee_{i=1}^\infty \theta_i$. Other connectives can be reduced to these. The converse is also true: to each $a \in \mathcal{F}$ there is a $\theta \in \mathcal{S}^\infty\mathcal{L}$ which is satisfied by just the truth functions in a. If a is a cylinder with places $i_1, \ldots, i_k$ fixed to T and places $j_1, \ldots, j_l$

[18] See [Karp, 1964].

[19] In the sense that $\models \neg(\theta_i \wedge \theta_j)$ for all i and j.

[20] [Billingsley, 1979], p. 27.

fixed to F, then $c_{i_1} \wedge \ldots \wedge c_{i_k} \wedge \neg c_{j_1} \wedge \ldots \wedge \neg c_{j_l}$ is the required sentence, and a countable union of cylinders requires a countable disjunction of such sentences. Now θ and ϕ both correspond to the same $a \in \mathcal{F}$ iff they are satisfied by the same set of truth functions, i.e. $\theta \equiv \phi$. Thus we have a bijection between $\mathcal{S}^{\infty}\mathcal{L}/\equiv$ and $\mathcal{F}$.

Such an infinitary propositional system is a useful formalism because it clarifies the link between probability on sentences and probability on events. However, many are cautious about using infinitary systems, either because they find the concept of infinitary sentences counterintuitive or for technical reasons — for example only a weak version of the completeness theorem can be proved. Interestingly, if we restrict attention to a finitary language and define probability accordingly then no probabilistic information is lost, in the sense that there are no more probability measures on an infinitary language than on the finitary language which it extends. Let $\mathcal{SL}$ denote the finitary sentences (involving the same countable set of propositional variables $\mathcal{L}$, but no countable disjunctions or conjunctions). Let $p : \mathcal{SL} \longrightarrow \mathbf{R}$ be a *(finitary) probability measure* if (for all $\theta, \phi \in \mathcal{SL}$)

F1: $0 \leq p(\theta) \leq 1$;

F2: $p(\theta \wedge \neg\theta) = 0$ and $p(\theta \vee \neg\theta) = 1$;

F3: if θ and ϕ are mutually exclusive then $p(\theta \vee \phi) = p(\theta) + p(\phi)$.

Define *conditional (finitary) probability* by:

FC: if $p(\theta) > 0$ then $p(\phi|\theta) = p(\theta \wedge \phi)/p(\theta)$.

Then a finitary probability measure on $\mathcal{SL}$ determines a unique probability measure on the infinitary extension $\mathcal{S}^{\infty}\mathcal{L}$.

The reasoning behind this fact is as follows. $\mathcal{SL}/\equiv$ is isomorphic to the field of cylinders of truth functions considered above. A finitary probability measure on $\mathcal{SL}$ then induces a finitely-additive probability measure on the field of cylinders. But any such measure on the field of cylinders must be countably additive,[21] and can therefore be uniquely extended to a (countably additive) probability measure on the generated σ-field $\mathcal{F}$.[22] By the isomorphism between $\mathcal{F}$ and $\mathcal{S}^{\infty}\mathcal{L}/\equiv$ this corresponds to a probability measure in the infinitary system.

One could of course define probability over first-order predicate sentences. But if we define existential and universal quantification in terms of denumerable disjunction and conjunction respectively, and the first-order language has countably many atomic sentences, then for the purposes of defining probability the first-order language can be thought of as an infinitary propositional language over the atomic sentences. Thus in this situation the extra

[21] [Billingsley, 1979] theorem 2.3.
[22] [Billingsley, 1979] theorem 3.1.

expressive power of the predicate calculus is linguistic rather than probabilistic. Coupling this reduction with the reduction to a finitary propositional language, we can uniquely extend a finitary probability measure over the finitary propositional language to a probability measure over the first-order language.[23]

In sum, an infinitary propositional system is a useful formalism, not only because of the link with the mathematical definition of probability, but also because it acts as a good half-way house between more familiar finitary propositional systems and first-order predicate systems.

2.1 Partial entailment

Having seen how probability can be defined on sentences, we are now in a position to examine the concept of a probability logic which generalises or modifies classical logic. Such a system can be classified according to its interpretation of the logical implication operator $\models$. Perhaps the most obvious thing to try first is a generalisation of entailment $\models$ to partial entailment $\models_x$, where a set Θ of sentences partially entails sentence ϕ to degree x, $\Theta \models_x \phi$, if and only if $p(\phi | \bigwedge \Theta) = x$. Under such a view classical entailment is the case where $x = 1$. If Θ is empty we get a concept of degree of logical truth or degree of truth which corresponds to unconditional probability. There have been some well-known proponents of this kind of view, as we shall see now. In the following subsection we will examine the viability of the partial entailment approach.

One may be able to attribute the germs of such a logical approach to probability to some of the pioneers of probability — for example Boole permitted probabilities defined on propositions, although his logical foundations of probability really amounted to a mathematical calculus of probability derived logically, rather than a generalisation of logic.[24] However it wasn't until the developments and interest in formal logic at the end of the 19^{th} and beginning of the 20^{th} centuries that the view of probability as logic began in earnest.

Frege, Peano and Russell's contributions to formal logic motivated Łukasiewicz' innovative theory.[25] Instead of defining probability over sentences as we do above, Łukasiewicz defines probability over *indefinite propositions*, formulae which contain free variables. An indefinite proposition is true or false if true or false respectively for all substitutions. Further, '*By the truth value of an indefinite proposition I mean the ratio between the number of values of the variables for which the proposition yields true judgements*

[23] See [Gaifman, 1964], [Scott and Kraus, 1966], [Paris, 1994] chapter 11, or [Roeper and Leblanc, 1999] chapter 5 for further details concerning such reductions.

[24] See [Boole, 1854], [Boole, 1854b], [Boole, 1854c].

[25] [Łukasiewicz, 1913].

and the total number of values of the variables'.[26] Relative truth value is defined as conditional probability. Thus partial truth and partial entailment are given a specific interpretation.

Łukasiewicz rejects probability over events as being too restrictive in that it represents only the single-case. He distinguishes subjective and objective probability, but finds both interpretations unsatisfactory, subjective probability because it is too psychologistic, too subjective, and beliefs are unmeasurable (this was before the betting set-up had been introduced), and objective probability because determinism renders it redundant (this was before quantum mechanics) and because the principle of the excluded middle states that a proposition is objectively true or false at every time, precluding objective partial truth. Instead Łukasiewicz' interpretation is intended to be a coherent explication of Laplace's notoriously problematic *principle of indifference*, which defines a probability to be the ratio of the number of cases favourable to an event to the total number of possible cases, if we are indifferent as to which case will occur.[27] As he says,

> The interpretation of the essence of probability presented here might be called the *logical* theory of probability. According to this viewpoint, probability is only a property of propositions, i.e., of logical entities, and its explanation requires neither psychic processes nor the assumption of objective possibility. *Probability, as a purely logical concept, is a creative construction of the human mind, an instrument invented for the purpose of mastering those facts which cannot be interpreted by universally true judgements (laws of nature).*[28]

Keynes was another key player in the partial entailment tradition. He argued that probability generalises logic, measuring the degree to which an argument is conclusive. However he also allowed a subjective interpretation to his logical view of probability. In effect he proposed a probabilistic logic of rational belief.[29] Jeffreys too thought of probability as a generalisation of deductive logic, expressing support for an inference, given data.[30] In this respect his probability theory was a formalisation of inductive logic.[31]

Hempel closely studied this inductive relationship between evidence and hypothesis, deriving a qualitative logic of confirmation with a well-defined syntax and semantics: 'Confirmation as here conceived is a logical relationship between sentences, just as logical consequence is'.[32] Carnap rendered

[26] [Lukasiewicz, 1913] pg. 17.
[27] [Laplace, 1814].
[28] [Lukasiewicz, 1913] pg. 38.
[29] See [Keynes, 1921] section 1.1.
[30] [Jeffreys, 1931] section 2.0.
[31] [Jeffreys, 1939] section 1.2.
[32] [Hempel, 1945] pg. 24.

Hempel's theory quantitative by bringing probability into the logic. For Carnap probability was degree of confirmation. This was not cached out in terms of frequency (which he thought to be a valuable concept but quite different) or subjective degrees of belief (which he argued are too psychologistic), but given a distinct logical interpretation. The issue of confirmation is 'a logical question because, once a hypothesis is formulated by h and any possible evidence by e ..., the problem whether and how much h is confirmed by e is to be answered by a logical analysis of h and e and their relations. This question is not a question of facts in the sense that factual knowledge is required to find the answer'.[33]

Jaynes explicitly adopted a Keynesian position, arguing in favour of a logical partial entailment approach together with a subjective interpretation.[34] According to Jaynes' own theory, background knowledge imparts a fixed degree of plausibility (formally a conditional probability) to a proposition, and it is through principles like the principle of indifference, Laplace's rule of succession, the maximum entropy principle and symmetry constraints that we can identify the correct probability. The maximum entropy principle, for example, says that one should assign probabilities over the sentences of a finite language $\{c_1, \ldots, c_N\}$ in such a way that the entropy ($-\sum_\alpha p(\alpha)\log p(\alpha)$, where the α range over the atomic states $\pm c_1 \wedge \ldots \wedge \pm c_N$, where $+c_i$ is c_i and $-c_i$ is $\neg c_i$) is maximised subject to any constraints imposed by background knowledge.[35]

2.2 Uniqueness

In this section we shall see that the partial entailment view of probability logic is not an easy one to maintain. In the next section we will discuss an alternative approach.

There are inconclusive objections specific to particular approaches. Language dependence is a trap for theories which rely on the principle of indifference, although its generalisation, the maximum entropy principle does not appear to be subject to these problems.[36] Theories that allow a subjective degree of rational belief interpretation to probability as defined over sentences suffer from the problem of logical omniscience. Here an agent is assumed to give the same degree of belief to logically equivalent sentences, even though the equivalence may not be known — the agent must somehow know all logical facts in order to be rational, which is rather a tall order, especially considering the fact that many difficult unsolved problems

[33] [Carnap, 1950] p. 20.

[34] [Jaynes, 1998] chapter 1.

[35] See [Paris, 1994] for justifications of the maximum entropy principle and an idea of the logical issues that surround the Keynes-Jaynes type of position.

[36] See [Paris, 1994], [Paris and Vencovská, 1997], [Paris, 1999].

in mathematics are questions of logical implication or equivalence.[37] On the other hand it is possible to represent uncertainty about logical implication even if we do adopt a subjective interpretation, as we shall see in the final section.

However, there is also an important general problem. The partial entailment approach requires that the degree x of confirmation that the set Θ of sentences gives to sentence ϕ, be a logical fact, dependent only on Θ and ϕ. Given the probabilistic interpretation of $\models_x$ as conditional probability, and letting Θ and ϕ vary, this means that there is some unique, distinguished probability function p^* which determines degree of confirmation, $\Theta \models_x \phi \Leftrightarrow p^*(\phi|\Theta) = x$. However, there are good reasons to doubt whether such a function p^* can be uniquely determined.

Łukasiewicz recognised the uniqueness requirement: 'Although probability does not exist objectively, the probability calculus is not a science of subjective processes and has a thoroughly objective nature. Hence the essence of probability must be sought not in a relationship between propositions and psychic states, but in a relationship between propositions and objective facts.'[38] Keynes considered the degree of belief interpretation unnecessary in as much as uniqueness leaves no room for subjectivity.[39] Jeffreys was of a similar opinion: the degree of belief interpretation is an optional extra. As he says, 'If we like there is no harm in saying that a probability expresses a degree of reasonable belief.'[40] Carnap also realised that if rational degree of belief is uniquely determined then the subjective element is gratuitous and can be omitted.[41] He puts it thus:

> The characterisation of logic in terms of correct or rational or justified belief is just as right but not more enlightening than to say mineralogy tells us how to think correctly about minerals. The reference to thinking may just as well be dropped in both cases. Then we say simply: mineralogy makes statements about minerals, and logic makes statements about logical relations. The activity in any field of knowledge involves, of course, thinking. But this does not mean that thinking belongs to the subject matter of all fields. It belongs to the subject matter of psychology but not to that of logic any more than to that of mineralogy.[42]

Jaynes on the other hand was a strict subjectivist and he thought that it is wrong to think of subjective entities as objective features of the physical

[37] See [Williamson, 1999c].
[38] [Lukasiewicz, 1913] p. 37.
[39] [Keynes, 1921] section 1.2.
[40] [Jeffreys, 1931] p. 22.
[41] [Carnap, 1950] section 2.11.
[42] [Carnap, 1950] pp. 41-42.

world, even if they do not vary from individual to individual.[43] However he did recognise that the uniqueness requirement precludes room for subjective disagreement (note that Jaynes' rational agent X is a robot):

> When we apply probability theory as the normative extension of logic, our concern is not with the personal probabilities that different people might happen to have; but with the probabilities that they "ought to" have, in view of their information ....
> In other words, at the beginning of a problem our concern is not with anybody's personal opinions, but with specifying the *prior information* on which our robot's opinions are to be based, in the context of the current problem. The principles for assigning prior probabilities consistently by logical analysis of that prior information are for us an essential part of probability theory.[44]

And,

> Surely the most elementary requirement of consistency demands that two persons with the same relevant prior information should assign the same prior probabilities.[45]

Thus uniqueness is recognised as a condition for the partial entailment position.[46] However there are two types of problem with uniqueness.

First, there seem general situations where two agents' beliefs may be rational yet differ. There are statements whose truth depends on the agent's perspective, such as "I am Xenelda". "Bob is tall" is a vague and to some extent subjective statement. Carnap might dismiss such statements as unscientific, and he defended his position against other examples.[47] More seriously, there are doubts as to how and when the constraining principles that Jaynes appeals to should be applied. For example, the principle of indifference may be applied in different ways depending on how the equipossible outcomes are delineated, and the maximum entropy principle can be interpreted as saying that one should not take risks, in as much as one should accept bets that minimise worst-case expected loss, which may not always be a good strategy.[48]

[43]This he called the *mind projection fallacy*: see [Jaynes, 1990] and [Jaynes, 1998] chapter 2 and pages 218, 1614.

[44][Jaynes, 1998] Appendix A p. 5.

[45][Jaynes, 1968] 228.

[46]Note that those who adopt the logical approach (with the uniqueness requirement) together with a subjective interpretation of probability (with Bayesian conditionalisation) occasionally call themselves 'objective Bayesians'. There is a possible source of confusion here: this position is objective in the sense that rational belief does not vary from individual to individual, *not* in the sense that I have been using, where an interpretation is objective if it is directly physical, not to do with subjects and their beliefs.

[47][Carnap, 1950] sections 46-47.

[48]See [Grünwald, 2000].

Second, even if these problems are overcome, further problems arise on infinite domains. One can formulate constraining principles over languages involving a finite number of propositional variables, but it is generally not possible to extend them to a denumerable language, in particular a predicate calculus. Carnap himself searched for a unique probability function representing degree of confirmation but ended up with a continuum of probability functions. We saw in the last section that he viewed confirmation as a logical matter, not factual. However, if a function must be chosen from Carnap's continuum of inductive methods, this must be done either factually through calibration (for instance, seeing in practice which function agrees with frequency or leads to the best long-term betting strategy), or subjectively in an arbitrary fashion. Others have found that intuitively plausible rational constraints contradict each other on the infinite domain.[49]

One can see that there can be no unique most rational probability function on an infinite domain for the following reason.[50] Suppose we have a denumerable sequence $\theta_1, \theta_2, \ldots$ of mutually exclusive and exhaustive sentences. Agent X has no information about any of these sentences — how should she set her beliefs? We might want to generalise the principle of indifference to claim that each sentence should be given equal degree of belief. However, this degree of belief must be zero, since if $p(\theta_i) = \varepsilon > 0$ for all i then $\sum_{i=1}^{\infty} p(\theta_i)$ diverges, but by countable additivity $\sum_{i=1}^{\infty} p(\theta_i) = p(\bigvee_{i=1}^{\infty} \theta_i) = 1$, since the θ_i are exhaustive. Then if $\varepsilon = 0$, we get $\sum_{i=1}^{\infty} p(\theta_i) = 0$ which also contradicts countable additivity. Therefore the principle of indifference does not generalise, and some θ_i and θ_j must be awarded different probabilities. Since there is no information about either of these sentences, a belief function that swaps their probabilities is equally rational. There is no unique most rational probability function.

Thus the uniqueness requirement poses difficulties for the partial entailment approach. However, there are several possible responses. Firstly there are many ways in which the approach can be altered to abandon such a requirement. Second one can retain the uniqueness requirement if one drops the subjective or logical interpretation in favour of an objective interpretation of probability. We shall consider these options in order.

2.3 Generalised partial entailment

Dropping the uniqueness requirement yields a new notion of entailment $\models_A$, where $\Theta \models_A \phi$ iff $A = \{x : x = p(\phi | \bigwedge \Theta) \text{ for some } p\}$.[51] This concept is a great deal weaker, for if Θ and ϕ are logically unrelated then $A = [0, 1]$ and

[49] [Wilmers *et al.*, 1999].

[50] [Williamson, 1999].

[51] [de Finetti, 1970] section 3.10 shows that A will always be a singleton or a closed interval.

the entailment relation says nothing. However, one can bolster the concept by limiting the probability functions under consideration when constructing the set A. Consider for example a subjective interpretation. Here the probability functions are rational belief functions of agents with knowledge Θ. Many subjectivists would accept the weak generalisation of partial entailment, but a subjectivist like Jaynes would apply extra principles, like maximum entropy, to narrow down the range of belief functions considered rational given the background knowledge. We saw in the last section that this need not narrow down A to a unique value, but it may nevertheless considerably strengthen the partial entailment relation.

Howson is a subjectivist of the former ilk. Degrees of belief are restricted to be probabilistic, but no further constraints are imposed (apart from the qualified use of Bayesian conditionalisation). Howson develops a logical approach to probability, based on the notion of consistency, which is defined by a set of axioms concerning fair bets. These are used to show that an assignment of degrees of belief is consistent if and only if it is the restriction of some probability measure.[52] We generate an entailment operator from this notion of consistency by appealing to Howson's identification of models with probability measures: an assignment of degrees of belief is consistent if and only if it has (is the restriction of) a model. Then assignment q entails assignment r iff all probability measures satisfying q, satisfy r. That is, q entails r iff r is the restriction of any probability measure p that extends q. If assignment q is of the form $q(\theta_1) = y_1, q(\theta_2) = y_2, \ldots$ and r is $r(\phi) = x$, this becomes: q entails r iff for all probability measures p, if $p(\theta_1) = y_1, p(\theta_2) = y_2, \ldots$ then $p(\phi) = x$. I shall call this formulation *probabilistic entailment*, and write $\theta_1/y_1, \theta_2/y_2, \ldots \models_x \phi$. Probabilistic entailment is just logical entailment via the axioms of probability. We then get the following connection with the earlier version of partial entailment: $\theta \models_x \phi \Leftrightarrow \forall p, p(\phi|\theta) = x \Leftrightarrow \forall p, p(\theta \wedge \phi) = xp(\theta) \Leftrightarrow \theta/y \models_{xy} \theta \wedge \phi$. Of course depending on the θ_i and ϕ there will often be no unique x such that $\theta_1/y_1, \theta_2/y_2, \ldots \models_x \phi$, in which case we can only say that $\theta_1/y_1, \theta_2/y_2, \ldots \models_A \phi$ for some set $A \subseteq [0,1]$. Then $\theta \models_A \phi \Leftrightarrow \theta/y \models_{Ay} \theta \wedge \phi$, where $Ay = \{xy : x \in A\}$. Thus Howson's proposal induces probabilistic entailment which is closely related to weak partial entailment.

Adams interprets probabilities as degrees of belief, but relies also on frequencies, since he argues that degrees of belief must approximate frequencies in order to be useful for practical reasoning.[53] Adams extends propositional logic by adding a new non-truth-functional conditional connective $\Rightarrow$, which can only link formulae which have the usual connectives in them (so one can have $\theta \Rightarrow (\phi \vee \psi)$ but not $\theta \Rightarrow (\phi \Rightarrow \psi)$)

[52] See [Howson, 2000]. The consistency condition is a normative constraint just as coherence is under the standard Dutch book foundations of subjective probability, which Howson rejects.

[53] [Adams, 1998] chapter 9 and appendix 1.

and for which $p(\theta \Rightarrow \phi) = p(\phi|\theta)$. Adams provides a probabilistic semantics: an inference is valid iff the uncertainty of its conclusion cannot exceed the sum of the uncertainties of its premises, where Adams defines uncertainty as $u(\theta) = 1 - p(\theta)$. It turns out[54] that an argument is valid in Adams' sense if and only if the premises entail the conclusion thus: $\forall\delta \in (0,1), \exists\varepsilon \in (0,1), [p(\theta_1) \geq 1-\varepsilon, \ldots, p(\theta_n) \geq 1-\varepsilon \Rightarrow p(\phi) \geq 1-\delta]$, which I will write as $\theta_1, \ldots, \theta_n \models_* \phi$ and call $*$-entailment. Note that $*$-entailment coincides with classical entailment on arguments which do not involve the connective $\Rightarrow$.[55]

2.4 Objective interpretations

While Howson and Adams are subjectivists, objective interpretations are also represented in the probability logic literature. Objective chance interpretations often appeal to a possible-world semantics for entailment. Above I equated the probability of a sentence with the probability of the set of truth functions that satisfy the sentence. One can think of this as the probability that the truth function corresponding to the actual world is in this set of truth functions, as the probability that the actual world is in the set of worlds at which the sentence is true, or as the proportion of possible worlds taken up with worlds at which the sentence is true. Nilsson's probability logic is based on this thought and employs the probabilistic entailment relation considered above.[56] One can also think of partial entailment in terms of possible worlds if one thinks of $p(\phi|\theta)$ as the proportion of θ-worlds in which ϕ is also true.[57] In this way one can retain the original concept of partial entailment, $\Theta \models_x \phi$ iff $x = p^*(\phi| \bigwedge \Theta)$, together with its uniqueness requirement, by interpreting p^* objectively.

Łukasiewicz gave the probability of an open sentence a logical interpretation, but it can also be given a frequency interpretation, in which the probability of $P(x)$ is the frequency of P. Coupling this with probabilities for sentences and a possible-worlds semantics, one can claim that the probability of a formula is the average frequency, with the average taken over classes of possible worlds and weighted by the chance of our world being in such a class. This is essentially the interpretation suggested by Loś and refined by Fenstad.[58]

---

[54] [Adams, 1998] section 7.1.

[55] $*$-entailment can also be used as a semantics for non-monotonic logic, if the numbers δ and ε are taken to be infinitesimals. See [Pearl, 1988] in this respect. Note however that non-standard probability measures lead to a range of conceptual problems, including failure of the archimedean property and difficulties to do with interpreting and measuring infinitesimal probabilities.

[56] [Nilsson, 1986].

[57] See [Adams, 1998] chapter 8.

[58] [Loś, 1963], [Fenstad, 1967].

Halpern adopts a similar approach, but instead of mixing the frequency distribution at a world with the chance distribution of the worlds, he makes a sharp distinction between the interpretation of the probabilities of open and closed formulae. The probabilities of open formulae are given a frequency interpretation, while the probabilities of sentences are given a distinct possible-world semantics.[59] In Halpern's theory, the probabilities of sentences are also thought of as subjective probabilities, but such an interpretation does not fit well with the possible-world semantics. It is much more intuitive and straightforward to apportion degrees of beliefs directly to sentences than to assess how likely our world is to be within a class of other possible worlds and then translate that probability to one over a sentence. Reformulating a problem in terms of possible worlds usually offers little or no extra insight for the price of a complicated and counter-intuitive ontology — possible worlds may be a necessary evil for chance theorists, but in my view they are best avoided by subjectivists.

3 PROBABILISTIC INFERENCE

I hope to have given an indication of the range of interpretations available, both to probability itself and to entailment in a probabilistic logic. In this section I will give a brief flavour of the approaches to inference using probability. In the following sections I shall argue for a reinterpretation of one of these approaches, and outline the ramifications for practical reasoning.

Probabilistic reasoning poses serious practical challenges. Consider the task of defining probability over a propositional language based on just a finite set of propositional variables $\mathcal{L} = \{c_1, \ldots, c_N\}$. To specify a probability measure (or finitary probability measure) over the sentences of this language, one must specify at least $2^N - 1$ probability values. For example, one can determine the probability measure from the values given to the 2^N atomic states, $p(\pm c_1 \wedge \ldots \wedge \pm c_N)$, and one of these is redundant since it can be determined by additivity from the others. Thus the *space complexity*, the amount of space required to store a probability measure, is exponential in the number of nodes. One can calculate the probabilities of other finitary sentences from the specified probabilities, but an exponential number of additions may be required. Therefore the *time complexity*, the amount of time required to perform a probabilistic inference, may severely restrict applications to practical reasoning. Likewise it may be practically difficult to check an assertion of entailment and even to check whether an assignment of probabilities over a set of statements is consistent (this is the *consistency problem*).

I will consider two types of strategy for overcoming these practical problems. First we shall look at probability logics and the approaches to infer-

[59][Halpern, 1990].

ence they yield. After this we shall look at a technique from artificial intelligence, namely Bayesian networks. Later I will argue that the Bayesian network approach can be integrated into a logic.

3.1 Probability logics

From a logical point of view, the first step towards efficient probabilistic inference is to provide a proof theory that is sound and complete with respect to the chosen entailment relation. Adams gives a sound and complete proof theory for his logic.[60] Likewise, probabilistic entailment can be given a sound and complete proof theory but only for the above finite language, not for example for a first order language involving an unbounded domain.[61] However, a proof theory is only a start. Finding a proof of a conclusion from its premises is often a prohibitively complex problem, even if a proof exists.

Nilsson gives a geometrical method for bounding the probabilities of sentences. Recall that for generalised probabilistic entailment, $\theta_1/y_1, \theta_2/y_2, \ldots \models_A \phi$ iff $\forall p$, if $p(\theta_1) = y_1, p(\theta_2) = y_2, \ldots$ then $p(\phi) \in A$. Nilsson gives a matrix technique for calculating the parameter A given a finite set of premises, their probabilities and the conclusion. Linear programming can be used to solve the problem, but as Nilsson acknowledges, computational complexity remains a serious difficulty.[62]

Logic programming aims to simplify the proof problem for classical logic by focussing on a more restricted language than the full predicate calculus, employing resolution as a single rule of inference, and treating the failure to find a proof of a ground atomic formula as a proof of its negation.[63] One hope is that by adding probabilities to logic programs and making use of the computational advantages offered by logic programming, one might perform probabilistic inference efficiently. Probabilistic Horn abduction, for instance, can be used to find the most likely hypothesis that explains a set of evidence.[64]

Another such system, stochastic logic programming, was devised to represent the bias of a machine learning program over the hypothesis and instance spaces,[65] but the formalism may also be applied to first-order probabilistic reasoning.[66] The basic idea here is that proofs of a goal (an implicitly existentially quantified open formula) are given a loglinear distribution parameterised by features of those proofs, and the probability of an instantiation of the goal is defined as the sum of the probabilities of the proofs that

[60] [Adams, 1998] chapter 7.
[61] [Halpern, 1990].
[62] See [Nilsson, 1986]. Nilsson also suggests a solution to the consistency problem.
[63] See [Nilsson and Maluszyński, 1990].
[64] [Poole, 1992].
[65] [Muggleton, 1995].
[66] [Cussens, 1999], [Cussens, 2000].

result in that instantiation. In a sense this progresses Łukasiewicz' original aim of defining probabilities on formulae based on their purely logical characteristics.

3.2 Bayesian networks

A Bayesian network consists of a directed acyclic graph, or *dag*, G over the propositional variables $c_1, \ldots, c_N$ together with a set of specifying probability values $S = \{p(c_i|d_i) : d_i$ is a state of the parents of c_i in G, $i = 1, \ldots, N\}$.[67] Now, under an independence assumption,[68] namely that given its parent states d_i, each node c_i in G is probabilistically independent of any state s of other nodes not containing the descendants of $c_i, p(c_i|d_i \wedge s) = p(c_i|d_i)$, a Bayesian network suffices to determine a probability measure p over the sentences $\mathcal{SL}$ of $\mathcal{L}$. This is determined from the probabilities of the atomic states, which are given by the formula $p(\pm c_1 \wedge \ldots \wedge \pm c_N) = \prod_{i=1}^{N} p(\pm c_i|d_i)$ where the d_i are the parent states consistent with the atomic state. Furthermore, any probability distribution on $\mathcal{SL}$ can be represented by some Bayesian network.[69]

In particular, if the c_i are *causal* variables, and the graph G represents the causal relations amongst them, with an arrow from c_i to c_j if c_i is a direct cause of c_j, then it is thought that G will automatically be acyclic, and the independence assumption will be a valid assumption.[70]

Bayesian networks offer the following advantages. First, if one deals just with Bayesian networks then there is no consistency problem, because any allocation of specifying values in $[0, 1]$ is consistent in that it is the restriction of some probability measure. Second, depending on the structure of the graph, the number of specifying probabilities may be relatively small. For example, if the number of parents of a node is bounded then the number of probabilities required to specify the measure is linear in N. Third, also depending on the structure of the graph, propagation techniques[71] can be employed which allow the quick calculation of conditional probabilities of the form $p(c_i|\alpha)$, where α is a state of other nodes. For example, if the graph is singly-connected (there is at most one path between two nodes) then propagation can be carried out in time linear in N. Thus while in the worst case (which occurs when the graph is *complete*, that is when there is an arrow between any two nodes) there is nothing to be gained by using a Bayesian network, if the graph is of a suitable structure then both the space

[67] If c_i has no parents, $p(c_i|d_i)$ is just $p(c_i)$.

[68] The Bayesian network independence assumption is often called the *Markov* or *Causal Markov* condition.

[69] See [Pearl, 1988] or [Neapolitan, 1990] for more on the formal properties of Bayesian networks.

[70] [Pearl, 1988], [Neapolitan, 1990].

[71] [Neapolitan, 1990].

and time complexity will be dramatically reduced. It is generally presumed that causal graphs are simple enough to offer satisfactory reductions in complexity.

4 BAYESIAN NETWORKS PROPERLY INTERPRETED

In this section and the next I would like to give some of my views as to the future directions of probability logic. I believe that Bayesian networks offer significant potential in this area, but that their use hinges on their interpretation. For a detailed account of the arguments in this section, see [Williamson, 2000].

The key issue is the interpretation of probability. Thus far in the Bayesian network literature the probability measure is either interpreted objectively, or subjectively but under the presumption that the subjective probabilities are estimates of objective probabilities. In my view, neither of these interpretations are viable.

The objective interpretation fails because the independence assumption need not hold, when assumed of causality with respect to objective probability. The independence assumption requires that any probabilistic dependency amongst the nodes in the network be accounted for by the causal relations within the network. This can be expressed more precisely by the *principle of the common cause*: if c_i and c_j are probabilistically independent and neither is a cause of the other, then they have one or more common causes and the dependency is *screened off* by the states d of the common causes, $p(c_i|d \wedge c_j) = p(c_i|d)$. However, the principle of the common cause can be shown to fail, for the simple reason that causal connection is not the only way that nodes can be rendered probabilistically dependent. They may be dependent because they have related meaning, or they may be logically or mathematically related, they may be related by non-causal physical laws, or by local or boundary conditions, or they may even be probabilistically dependent purely by accident. In any of these eventualities the Bayesian network independence assumption can fail.

On the other hand, if we view the Bayesian network as the knowledge of an agent X, consisting of her picture of causality and her degrees of belief, and those degrees of belief are assumed to be estimates of objective probabilities, then there is a further reason why the independence assumption might fail. X may simply not know about all the causal variables relevant to those in her language, or she may not know of all the causal relations linking the variables already in her language. One can show that if X's causal graph is incomplete in this sense then the independence assumption is very unlikely to hold.

If we opt for an unconstrained subjective interpretation, losing the link between subjective and objective probabilities, then there is no reason *at*

all to suppose the independence assumption might hold. For X may have any belief function she wishes, and only a few of those (a set of measure zero, to be precise) will satisfy any non-trivial independence assumption.

The only alternative, I claim, is to adopt a constrained subjective interpretation, along the lines of Jaynes' interpretation, by employing the maximum entropy principle.[72] The idea is that the components of the Bayesian network — the causal graph G and specified probabilities S — represent agent X's background knowledge, and the maximum entropy principle is used to derive the full probability measure p. The specified probabilities offer an immediate constraint on this process: the derived measure p must extend these specifiers. There is no immediate constraint imposed by the causal graph, and I propose the principle of *causal irrelevance* be used to invoke a constraint. This principle says that if a probability measure p is derived over $\mathcal{SL}$ via maximum entropy, next a new propositional variable c_{N+1} which is not a cause of any of the variables in $\mathcal{L}$ is added to $\mathcal{L}$ to give $\mathcal{L}^+$, and finally a new probability measure q over $\mathcal{SL}^+$ is derived, then q extends p, that is, the restriction $q|_{\mathcal{SL}} = p$. Intuitively there is no information that c_{N+1} is relevant to the other variables, so it should be considered irrelevant. It can be shown[73] that

E: under these constraints, the probability measure p derived by maximum entropy is the same as that derived by a Bayesian network under the independence assumption.

Thus if probability measure p is given this type of subjective interpretation, and is constrained by G and S, then the independence assumption holds and p can be represented by the Bayesian network on G and S.

5 NEW DIRECTIONS

5.1 A causal logic

Given the above interpretation of the components of a Bayesian network as background knowledge, we can define a logic, as follows. As before, we consider a language based on a set of *causal* propositional variables, and for practical reasons this set is finite. The components (causal graph G and probability specification S) of a Bayesian network together with a set of sentences Θ *partially entail* sentence ϕ to degree x, $G, S, \Theta \models_x \phi$, iff, given the information represented in G and S, an agent ought to believe ϕ to degree x, $p(\phi | \bigwedge \Theta) = x$. The maximum entropy principle is used to select the function p (uniquely, because the language is finite, and so we consider

[72]There are a finite number of propositional variables here, so we don't get problems of non-uniqueness.
[73][Williamson, 2000].

degree x rather than set A of degrees) and therefore the degree x of partial entailment. The value x can be determined from the Bayesian network formed on the components G and S. Such a calculation can be viewed as a *proof* of the entailment, and by equivalence [E], Bayesian network proofs are sound and complete with respect to this concept of partial entailment. The space and time complexity of a proof depend on the structure of the graph G, and the time complexity also depends on the form of Θ and ϕ.

Recall that propagation techniques can be used to quickly calculate $p(a_i|\alpha)$ where α is a state of some of the other variables. These techniques can also be used to calculate $p(\phi|\theta)$, as follows. First note that $p(\phi|\alpha)$ can be broken down into propagations. Write ϕ in disjunctive normal form as $\bigvee_{j=1}^{m} \alpha_j$ where each α_j is of the form $\bigwedge_{k=1}^{r} \pm c_{i_k}$ and the α_j are mutually exclusive. Now $p(\alpha_j|\alpha) = 0$ if α_j is inconsistent with α, otherwise $p(\alpha_j|\alpha) = p(\pm c_{i_1}|\pm c_{i_2} \wedge \ldots \wedge \pm c_{i_r} \wedge \alpha)p(\pm c_{i_2}|\pm c_{i_3} \wedge \ldots \wedge \pm c_{i_r} \wedge \alpha) \ldots p(\pm c_{i_r}|\alpha)$. Then $p(\phi|\alpha) = \sum_{j=1}^{m} p(\alpha_j|\alpha)$ and can thus be decomposed into propagations. Now $p(\phi|\theta)$ can also be decomposed into propagations, for if we write θ in disjunctive normal form as $\bigvee_{j=1}^{m} \alpha_j$ where each α_j is of the form $\bigwedge_{k=1}^{r} \pm c_{i_k}$ and the α_j are mutually exclusive, then

$$p(\phi|\theta) = \frac{p(\theta|\phi)p(\phi)}{p(\theta)} = \frac{\sum_{j=1}^{m} p(\alpha_j|\phi)p(\phi)}{\sum_{j=1}^{m} p(\alpha_j)} = \frac{\sum_{j=1}^{m} p(\phi|\alpha_j)p(\alpha_j)}{\sum_{j=1}^{m} p(\alpha_j)}.$$

Thus propagation techniques can be used to perform a proof, but the more logically complex the sentences involved in the partial entailment, the greater the time complexity of the proof.

Given such a causal logic, one can ask what should happen when an agent's background knowledge does not take the form of the components G and S of a Bayesian network. In particular, how should one derive a probability measure if not all of the required probability specifiers are available?

Garside [1996] and Rhodes [1999] have provided answers to this question for the special cases in which the causal graph has a tree or inverted tree structure. However, while they appeal to the maximum entropy principle they do not allow the causal knowledge to constrain the maximum entropy solution in any way. I advocate the principle of causal irrelevance as an extra constraint, and this gives a different solution to the problem. I shall indicate here how the combination of causal irrelevance and maximum entropy can be used to shed light on the issue of incomplete background knowledge.

Suppose X's background knowledge consists of a causal dag G with each arrow $c_i \rightsquigarrow c_j$ labelled by a weight $w \in [-1, 1]$, and each root node labelled by its probability. If the weight is positive, this signifies that $p(c_j|c_i \wedge d) \geq p(c_j|\neg c_i \wedge d) + w$ for each state d of the other parents of c_j. If it is negative then $p(c_j|c_i \wedge d) \leq p(c_j|\neg c_i \wedge d) + w$ for each such state d (intuitively in this case c_i prevents c_j). Using the techniques involved in the proof of the equivalence result [E], one can show that the

probability measure determined from such a labelled graph by the principles of causal irrelevance and maximum entropy is the same measure as that determined by the Bayesian network involving specifiers of the form $p(c_j|c_i \wedge d) = \frac{1+w}{2}, p(c_j|\neg c_i \wedge d) = \frac{1-w}{2}$. One may specialise further by considering labels of the form $w \in \{+, -\}$, with some $\varepsilon > 0$ given such that $c_i \rightsquigarrow^{+} c_j$ implies $p(c_j|c_i \wedge d) \geq p(c_j|\neg c_i \wedge d) + \varepsilon$ and $c_i \rightsquigarrow^{-} c_j$ implies $p(c_j|c_i \wedge d) \leq p(c_j|\neg c_i \wedge d) - \varepsilon$. We then get an equivalence between such a structure and a Bayesian network in which if $c_i \rightsquigarrow^{+} c_j$ then $p(c_j|\pm c_i \wedge d) = \frac{1\pm\varepsilon}{2}$, and similarly for prevention (and if root probabilities are omitted in the labelled graph they are specified to be $\frac{1}{2}$ in the Bayesian network). Thus we can define partial entailment $G, \Theta \models_x \phi$ where G is a labelled graph of one of the above varieties. This move not only extends the causal logic to situations where background knowledge is more limited, but can also significantly reduce the space complexity of the corresponding Bayesian network. In general, if we replace the values in the specification S by bounds on those values, we determine the specification S, and thereby a Bayesian network, by selecting the values within those bounds which maximise entropy. The task of finding the probability measure over the whole domain that maximises entropy is broken down into smaller and easier tasks, namely those of maximising the entropy of the individual specifying probabilities.

5.2 A proof logic

I claimed earlier that causal connection is only one type of link between variables, and that other links, such as logical relations, may also induce probabilistic dependencies. This observation motivates the hope that the Bayesian network toolkit may be applicable to these other types of links. I shall consider one such application here.

A logical proof of a sentence takes the form of an ordered list. Consider a propositional language with sentences $s, t, u, \ldots$ and the following proof of $s \to t, t \to u \vdash s \to u$, using the axiom system of [Mendelson, 1964] section 1.4:

1. $t \to u$ [hypothesis]
2. $s \to t$ [hypothesis]
3. $(s \to (t \to u)) \to ((s \to t) \to (s \to u))$ [axiom]
4. $(t \to u) \to (s \to (t \to u))$ [axiom]
5. $s \to (t \to u)$ [by 1, 4]
6. $(s \to t) \to (s \to u)$ [3, 5]
7. $s \to u$ [2, 6]

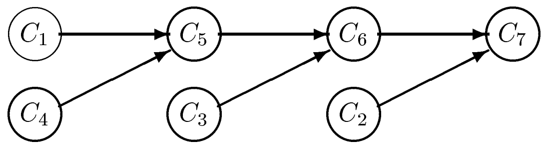

Figure 1. A proof dag.

The first important thing to note is that the ordering in a proof defines a directed acyclic graph. If we let c_i signify the sentence on line i, for $i = 1, \ldots, 7$, we get the dag in Figure 1.

By specifying degrees of belief in root nodes and conditional degrees of belief in other nodes given states of their parents, we can form a Bayesian network. These beliefs will depend on the meaning of the sentences, and in this example a specification might start like this: $S = \{p(c_1) = \frac{3}{4}, p(c_2) = \frac{1}{3}, p(c_3) = 1, p(c_4) = 1, p(c_5|c_1 \wedge c_4) = 1, p(c_5|c_1 \wedge \neg c_4) = \frac{1}{2}, \ldots\}$.

We can interpret the probability measure in various ways and if we assume that the probabilities are estimates of objective probabilities then, just as with causal Bayesian networks, we would not expect the independence assumption to hold unless (i) there are no dependencies due to non-logical relations amongst the variables and (ii) all the logical relations are included in the proof graph (but note that the graph will not necessarily remain acyclic if it includes all logical relations).

However the analogy with the causal case extends further, and if we adopt a subjective interpretation constrained by the maximum entropy principle, we would expect *proof irrelevance*, the analogue of the causal irrelevance condition, to hold. For if agent X learns of a new sentence c_8 that does not logically imply any of the others, her degrees of belief in the other sentences should intuitively not change.[74] Then the equivalence argument [E] can be used to justify equating X's belief function with the probability measure determined by the Bayesian network.

Finally, we can form a proof logic in the same way that we formed a causal logic above. While the parents of a node logically entail it, each node or set of nodes will also partially entail the others in the proof graph. The partial entailment relation can be proved, or the degree of partial entailment can be found, by using Bayesian network calculations. Thus two senses of proof are in play at once: a Bayesian network is used to prove a partial entailment, and the classical axiomatic method is used to prove the logical entailment on which the network is based.

[74]X learns of c_8 in the sense that she extends her language to include the new sentence, not in the sense of her learning the truth or falsity of c_8, which may well give her reason to change her other degrees of belief.

Why form beliefs about sentences in a proof? Because beliefs play a key role in proof planning. As Corfield demonstrates, subjective probability is important in mathematics: mathematicians regularly require degrees of belief in mathematical propositions, and these are apportioned according the mathematical and physical evidence available at the time.[75] The degrees of belief are instrumental in deciding whether to tackle a proof of a proposition (it may be worth tackling if the belief in the conclusion conditional on the premises is above a certain threshold) and in predicting from which area a possible proof will come (a conclusion may be more probable conditional on one set of premises or intermediary lemma than on another). To apply Bayesian proof networks to this type of problem we must recognise first that a parent of a node in the graph need not be one rule of inference away from the node. Just as a causal graph may represent causality on the macro-scale as well as the micro-scale, so too Bayesian proof networks may represent arguments involving large logical steps, as well as proofs like the one exemplified above. Second, just as in the causal case, an agent may not know of all the relevant logical factors or logical relations, and the proof network need not represent a complete proof — the proof graph will still be a dag. Thus the proof network includes the premises, conclusion and various other facts or conjectures which are considered relevant to the problem. The arrows in the network represent the flow of the proposed proof, and the probability specifiers represent the degrees of belief in each particular proposition, conditional on states of their parents. In this way a Bayesian proof network can be used to represent a belief function over a realistic mathematical problem, and to evaluate the conjectures and proposed proof paths.

Proof planning is not just important in mathematics. Whatever the domain and whatever the logic, if the proof theory relies on axiomatic deduction then finding a proof is likely to be a hard problem.[76] Automated theorem proving is now a large field of research, with applications ranging from software verification to robotics, but few systems tackle uncertainty in a fundamental way. Bayesian proof networks offer a practical opportunity for doing so. Moreover, proof planning is not the only application of such networks. Any domain of reasoning which requires the assessment of logically complex and connected variables may benefit from the approach, just as causal Bayesian networks are important for any kind of reasoning with causal variables, for example diagnosis, prediction and causal explanation.

[75] [Corfield, 2000].
[76] [Bundy, 1999], [Bundy, 2000].

6 SUMMARY

The choice of an appropriate interpretation of probability is of key importance, especially when it comes to combining probability with logic. Many logical approaches to probability ultimately fail from a philosophical point of view because of the uniqueness problem on infinite domains. However, probability logic can be sustained either by using a more general conception of partial entailment, by appealing to objective probability, or by sticking to finite domains. Bayesian networks take this latter approach and offer a way out of the practical problems that face a logic which incorporates probability, but again we must be careful about how we interpret probability. Probability for logic is best interpreted subjectively, using constraining principles like Jaynes' maximum entropy principle, if we wants to tap the power of Bayesian networks.

This power lies not only at the computational level, but also at the level of applications: Bayesian networks lead rather naturally to a causal logic and a proof logic. The causal logic may be applied to the problem posed at the beginning of this chapter by constructing a causal graph and probability specification (or bounds on these probability values), and using the associated Bayesian network to calculate the probability of X's house being burgled, given that her window is open and she hears an alarm. The proof logic may be applied to proof planning. Further extensions may be possible, and in the future we may expect some sort of integration between these extensions, either horizontally, where languages may involve mixtures of causal and logically complex variables for instance, or vertically where one logic is applied to a domain of causal variables and then a metalogic is applied to the logically complex sentences formed over that domain.

ACKNOWLEDGEMENTS

Thanks to David Corfield and Donald Gillies for useful comments, and to the Arts and Humanities Research Board for funding this research.

BIBLIOGRAPHY

[Adams, 1998] Ernest W. Adams. *A Primer of Probability Logic*, Stanford: CSLI Publications.

[Billingsley, 1979] Patrick Billingsley. *Probability and Measure*, John Wiley & Sons, third edition 1995.

[Boole, 1854] George Boole. *An Investigation of the Laws of Thought on which are Founded the Mathematical Theories of Logic and Probabilities*, London: Walton and Maberly.

[Boole, 1854b] George Boole. *Sketch of a Theory and Method of Probabilities Founded upon the Calculus of Logic*, in Studies in Logic and Probability, Watts & Co 1952, pages 141–166.

[Boole, 1854c] George Boole. *On a General Method in the Theory of Probabilities*, in Studies in Logic and Probability, Watts & Co 1952, pages 291–307.

[Bundy, 1999] Alan Bundy. A survey of automated deduction, Edinburgh Artificial Intelligence Research Paper 950.

[Bundy, 2000] Alan Bundy. A critique of proof planning, in [Kakas and Sadri, 2000].

[Carnap, 1950] Rudolf Carnap. *Logical Foundations of Probability*, Routledge and Kegan Paul Ltd.

[Corfield, 2000] David Corfield. Bayesianism in mathematics, in [Corfield and Williamson, 2000].

[Corfield and Williamson, 2000] David Corfield and Jon Williamson, eds. *Foundations of Bayesianism*. Kluwer Academic Publishers, 2001.

[Cussens, 1999] James Cussens. Loglinear models for first-order probabilistic reasoning. In *Proceedings of the 15th Annual Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, pages 126–133.

[Cussens, 2000] James Cussens. Stochastic logic programs: sampling, inference and applications. In *Proceedings of the 16th Annual Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, pages 115–122.

[Fenstad, 1967] J.E. Fenstad. Representations of probabilities defined on first order languages, in John N. Crossley, ed. *Sets, Models and Recursion Theory: Proceedings of the Summer School in Mathematical Logic and Tenth Logic Colloquium*, pages 156–172.

[Field, 1977] Hartry H. Field. Logic, meaning and conceptual role. *Journal of Philosophy*, **74**, 379–409, 1977.

[de Finetti, 1937] Bruno de Finetti. Foresight. Its logical laws, its subjective sources. In [Kyburg and Smokler, 1964], pages 93–158.

[de Finetti, 1970] Bruno de Finetti. 'Theory of probability', Wiley, 1974.

[Gaifman, 1964] H. Gaifman. Concerning measures in first order calculi. *Israel Journal of Mathematics* **2**, 1–18, 1964.

[Garside and Rhodes, 1996] Gerald R. Garside and Paul C. Rhodes. Computing marginal probabilities in causal multiway trees given incomplete information. *Knowledge-Based Systems*, **9**, 315–327, 1996.

[Garside *et al.*, 1999] G.R. Garside, P.C. Rhodes and D.E. Holmes. The efficient estimation of missing information in causal inverted multiway trees. *Knowledge-Based Systems*, **12**, 101–111, 1999.

[Grünwald, 2000] Peter Grünwald. Maximum entropy and the glasses you are looking through. In *Proceedings of the 16th conference of Uncertainty in Artificial Intelligence*, Stanford, 2000.

[Halpern, 1990] Joseph Y. Halpern. An analysis of first-order logics of probability. *Artificial Intelligence*, **46**, 311–350, 1990.

[Hempel, 1945] Carl G. Hempel. Studies in the logic of confirmation. In *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science*, New York: The Free Press 1965, 1970, pages 3–51.

[Howson, 1995] Colin Howson. Theories of probability. *British Journal for the Philosophy of Science*, **46**, 1–32, 1995.

[Howson, 2000] Colin Howson. The logical basis of uncertainty, in [Corfield and Williamson, 2000].

[Jaynes, 1968] E. Jaynes. Prior probabilities. In *IEEE Transactions Systems Science and Cybernetics*, *SSC-4*(3), 227, 1968.

[Jaynes, 1990] E.T. Jaynes. Probability theory as logic, in Paul F. Fougere, ed. *Maximum Entropy and Bayesian Methods*, Kluwer, 1990.

[Jaynes, 1998] E.T. Jaynes. Probability theory: the logic of science, http://bayes.wustl.edu/etj/prob.html.

[Jeffreys, 1931] Harold Jeffreys. *Scientific Inference*, Cambridge: Cambridge University Press, second edition 1957.

[Jeffreys, 1939] Harold Jeffreys. *Theory of Probability*, Oxford: Clarendon Press, third edition 1961.

[Kakas and Sadri, 2000] A. Kakas and F. Sadri. *Essays in Honour of Robert Kowalski*, to appear.

[Karp, 1964] C.R. Karp. *Languages with Expressions of Infinite Length*, North-Holland, 1964.

[Keynes, 1921] John Maynard Keynes. *A Treatise on Probability*, London: Macmillan, 1948.

[Kolmogorov, 1933] A.N. Kolmogorov. *The Foundations of the Theory of Probability*, New York: Chelsea Publishing Company, 1950.

[Kyburg and Smokler, 1964] H.E. Kyburg and H.E. Smokler, eds. *Studies in Subjective Probability*, New York: John Wiley, 1964.

[Laplace, 1814] Pierre Simon-Marquis de Laplace. *A Philosophical Essay on Probabilities*, New York: Dover, 1951.

[Lewis, 1980] David K. Lewis. A subjectivist's guide to objective chance, in [Lewis, 1986], pages 83–132, 1980.

[Lewis, 1986] David K. Lewis. *Philosophical Papers Volume II*, New York: Oxford University Press, 1986.

[Loś, 1963] J. Loś. Remarks on the foundations of probability. In *Proceedings of the 1962 International Congress of Mathematicians*, pages 225–229, 1963.

[Łukasiewicz, 1913] Jan Łukasiewicz. Logical foundations of probability theory. In L. Borkowski, ed. *Jan Łukasiewicz Selected Works*, Amsterdam: North-Holland 1970, pages 16–63.

[Mellor, 1971] David H. Mellor. *The Matter of Chance*, Cambridge: Cambridge University Press, 1971.

[Mendelson, 1964] Elliott Mendelson. *Introduction to Mathematical Logic*, Chapman and Hall, fourth edition 1997.

[von Mises, 1928] Richard von Mises. *Probability, Statistics and Truth*, Allen and Unwin, 2nd edition 1957.

[von Mises, 1964] Richard von Mises. *Mathematical Theory of Probability and Statistics*, Academic Press, 1964.

[Muggleton, 1995] Stephen Muggleton. Stochastic logic programs, in L. De Raedt, ed. *Advances in Inductive Logic Programming*, IOS Press, 1995.

[Neapolitan, 1990] Richard E. Neapolitan. *Probabilistic Reasoning in Expert Systems: Theory and Algorithms*, New York: Wiley, 1990.

[Neapolitan, 1992] Richard E. Neapolitan. A limiting frequency approach to probability based on the weak law of large numbers. *Philosophy of Science*, **59**, 389–407, 1992.

[Nilsson, 1986] Nils J. Nilsson. Probabilistic logic, *Artificial Intelligence*, **28**, 71–87, 1986.

[Nilsson and Maluszyński, 1990] Ulf Nilsson and Jan Maluszyński. *Logic, Programming and Prolog*, Chichester: John Wiley andSons, second edition, 1995.

[Paris, 1994] Jeff Paris. *The Uncertain Reasoner's Companion*, Cambridge: Cambridge University Press, 1994.

[Paris, 1999] Jeff Paris. Common sense and maximum entropy, *Synthese*, **117**, 73–93, 1999.

[Paris and Vencovská, 1997] Jeff Paris and Alena Vencovská. In defense of the maximum entropy inference process, *International Journal of Automated Reasoning*, **17**, 77–103, 1997.

[Pearl, 1988] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo, California: Morgan Kaufmann, 1988.

[Peirce, 1910] Charles Sanders Peirce. Notes on the doctrine of chances. In *Collected papers, Vol 2*, Harvard University Press 1932, page 404.

[von Plato, 1994] Jan von Plato. *Creating Modern Probability: Its Mathematics, Physics and Philosophy in Historical Perspective*, Cambridge: Cambridge University Press, 1994.

[Poole, 1992] D. Poole. Logic programming, abduction and probability. In *Proceedings, International Conference on Fifth Generation Computer Systems 1992*, vol 2, page 530, 1992.

[Popper, 1934] Karl R. Popper. *The Logic of Scientific Discovery*, London: Hutchinson, 1974.

[Popper, 1972] Karl R. Popper. *Objective Knowledge: An Evolutionary Approach*, Revised 1979, Oxford: Oxford University Press.

[Popper, 1983] Karl R. Popper. *Realism and the Aim of Science*, Hutchinson, 1983.

[Ramsey, 1926] Frank Plumpton Ramsey. Truth and probability. In [Kyburg and Smokler, 1964], pages 61–92, 1926.

[Roeper and Leblanc, 1999] P. Roeper and H. Leblanc. *Probability Theory and Probability Logic*, University of Toronto Press, 1999.

[Scott and Kraus, 1966] Dana Scott and Peter Kraus. Assigning probabilities to logical formulas. In Jaakko Hintikka and Patrick Suppes, eds. *Aspects of Inductive Logic*, Amsterdam: North-Holland, pages 219–264, 1966.

[Williamson, 1999] Jon Williamson. Countable additivity and subjective probability, *British Journal for the Philosophy of Science*, **50**(3), 401–416, 1999.

[Williamson, 1999b] Jon Williamson. The actual frequency interpretation of probability, philosophy.ai report pai_jw_99_b, http://www.kcl.ac.uk/philosophy.ai.

[Williamson, 1999c] Jon Williamson. Logical omniscience and rational belief, philosophy.ai report pai_jw_99_e, http://www.kcl.ac.uk/philosophy.ai.

[Williamson, 2000] Jon Williamson. Foundations for Bayesian networks. In [Corfield and Williamson, 2000], 2000.

[Wilmers *et al.*, 1999] G.M. Wilmers, M.J. Hill and J.B. Paris. Some observations on induction in predicate probabilistic reasoning, Department of Mathematics, Manchester University, 1999.

LUÍS MONIZ PEREIRA

PHILOSOPHICAL INCIDENCE OF LOGIC PROGRAMMING

1 INTRODUCTION

We begin by considering the general way in which artificial intelligence (*AI*) impinges on philosophical logic, by examining the requirements posed on logic by knowledge representation and reasoning issues which AI has addressed. We then outline some of AI's contributions, most especially via the field of logic programming (*LP*), to more dynamic forms of logic, in order to deal with knowledge in flux, namely: incomplete and contradictory information; hypothesis making through abduction; argumentation; diagnosis and debugging; updating; and learning. Along the way we delve into implications for the philosophy of knowledge.

Our motivation is to entice the philosophically inclined reader, as well as the *LP* veteran, to the new territories recently mapped out in the logical continent by ongoing research in *AI* through LP, by providing a *personal* bird's eye view of the region. The technical parts of the exposition aim at an intuitive capturing of the general issues by rather simple examples, and of how they are tackled and overcome. Detailed formal treatment is deferred to the abundant references.

The structure of the chapter is as follows: The next section recapitulates and emphasizes the unique rôle of logic as a human knowledge representation and reasoning tool. The section that follows prolongs the one before to stress the benefit of mechanizing logic on the computer. Subsequently, the different encroachments of *AI* on logic are listed, and the last one, a framework whose fresh import for indexphilosophy philosophy is underscored, is selected for further analysis. In the sequel, some of the LP tools bearing on the proposed framework are brought out. Afterwards, the use of such tools for diagnosing a logical theory is outlined. In the section following that a dynamical logic framework is sketched, and then elaborated upon by summarily describing the *MENTAL* project. Finally, there are Concluding Remarks and, at the end, a challenge for philosophers.

2 EVOLUTION AND REASONING

Evolution has provided humans with symbolic thought, and symbolic language communication abilities.

Objective common knowledge requires thought to follow abstract, content independent rules of reasoning and argumentation, which must not be entirely subjective, on pain of making precise communication and collective rational endeavour impossible. Such rules have become ingrained in human thought, and hold an enormous joint survival value [Deacon, 1997; Donald, 1991; Pinker, 1997].

In human cognitive evolution, both mimetic knowledge (such as that inherent in reality-simulating maps and models), and imitation knowledge (such as that present in ritual observation, or in artifact reproduction), were essential first steps towards socially situated, joint rule following behaviour, required by, say, hunting plans. Subsequently, throughout human cultures, abstract rule following social games emerged.

Game rules encapsulate concrete situation defining patterns, and concrete situation-action-situation causal sequencing, which mirrors causality-obeying physical reality. From games, further abstraction ensued, and there finally emerged the notions of situation-defining concepts, of general rules of thought and their chaining, and of legitimate argument and counter-argument moves. Together they compose a cognitive meta-game [Holland, 1998].

The pervasiveness of informal logic for capturing knowledge and for reasoning, a veritable *lingua franca* across human languages and cultures, rests on its ability to actually foster rational understanding and common objectivity. Crucially, objective knowledge evolution dynamics, whether individual or plural, follows ratiocination patterns and laws.

Moreover, and more recently, the very same rules of reasoning can and are employed to reason about reasoning. Though some reasoning methods are well known, some are still unconscious but, like the rules of grammar, can be discovered through research. What is more, new reasoning methods can and have been invented and perfected throughout human history. Examples of these are transfinite induction, *reductio ad absurdum* (proof by contradiction), recursion, abduction, and contradiction removal, to name but a few.

New purported reasoning methods may be disputed, just like any specific train of reasoning can. But reasoning can only be disputed by further reasoning, if any consensus is to be found! (cf. [Nagel, 1997]). Some argue that scientific and philosophical discussion is necessarily tantamount to a culture sensitive, and culturally relative, persuasive informal *ad hoc* argumentation, allied to anything goes rhetoric [Gross and Levitt, 1994]. They ignore that argumentation is just another form of reasoning which has itself been made the subject of logical formalization, and are oblivious to the fact that rhetoric may be fine for preachers, but is not conducive to the two-sided communication required to reach common agreement in the all rigorous scientific praxis that lead to cumulative knowledge.

3 LOGIC AND THE COMPUTER

Logic, we have seen, has arisen as the content independent formulation of the Laws of Thought. Predicates express whatever conceptual relations we may wish, about the external and internal worlds. These predicate building blocks can then be combined into formulas, by means of quantifiers and logical connectives, and can be manipulated according to inference rules characterizing diverse forms of reasoning.

Logic is capable of articulating an extensional behaviouristic view of predicates, seen as black boxes or input/output relations, simultaneously with an intensional view of predicates, seen as functionally decomposable into other predicates, like black boxes wired inside black boxes [Monteiro and Pereira, 1970].

Since the language of logic is symbolic and its rules content-independent, its workings can be specified by general, abstract, rule following procedures. These, in turn, can be programmed on the computer. Through their mechanization, logical theories and reasoning forms attain, for the first time, an *in vitro* existence, an ability and availability for repeatable execution, independently of any human mind hardware support [Pereira, 1988].

The two views above, extensional and intensional, are reconciled in the computer by insisting that the declarative semantics, the "what" is to be computed, be equivalent to the procedural computational semantics, the "how" it is computed. This is the cornerstone of the logic programming paradigm, of which more later [1].

4 PHILOSOPHY AND AI

Philosophy has been a cradle for rational thought throughout human intellectual history. Its study of language fostered the refinement of logic as an abstraction of natural language and as a tool of linguistic inquiry. The investigations on the foundations of mathematics assigned to logic a meta-instrumental rôle that would deploy it as the general purpose symbol manipulating tool *par excellence*. Computer science as well has adopted logic as its general foundational tool, while *AI* has made viable the proposition of turning it into a *bona fide* computer programming language.

At the same time, *AI* has developed logic beyond the confines of monotonic cumulativity, typical of the precise, complete, endurable, condensed, and closed mathematical domains, in order to open it up to the non-monotonic real world domain of imprecise, incomplete, contradictory, arguable, revisable, distributed, and evolving knowledge. In short, *AI* has added dynamics to erstwhile statics.

In *AI*, as with any science, common logic has an important rôle in it. Indeed, any science, implicitly or explicitly, presses logic and reasoning into its service, to build up arguments, counter-arguments, and to settle disputes. But a greater rôle can be expected of *AI* in logic!

- First, *AI* means to mechanize logic, it being such a core tool for so many rational activities.

- Second, *AI* intends to make explicit, and well-defined, the unconscious logic we use, and put it to objective test by automating it on the computer.

[1] Henceforth, Science and Technology, with the support of Computer Science and most especially of AI, will more and more see automated, in symbiosis with Man, their own processes of creation, in positive retroaction!

- Third, *AI* employs logic as a generic language of communication, of knowledge and of procedures, between humans and computers, and among computers themselves.
- Fourth, in *AI*, even when procedures and devices are not implemented using logic directly, logic can play the rôle of a precise specification language for their requirements, and of a formalism with which to study their semantical properties.
- Fifth, *AI* has contributed significantly to the phrasing and examination of the problem of identifying limits to symbolic reasoning embodied in computers, and whether these limits apply to humans as well.
- Sixth rôle, *AI* has helped researchers explore new reasoning issues and methods, and to combine disparate reasoning modalities into a uniform unified framework, so as to deal with incomplete, imprecise, contradictory, and changing information.

This last rôle, in contrast with its predecessor, much pampered by philosophers, has hardly caught their attention yet. The latter-day epigenesis of logic it announces, and its computational scaffolding, point to an innovative rôle for the logic enterprise. To wit, that of supporting, justifying, and mechanizing the dynamics of the abstract epistemic and argumentative processes underlying scientific endeavour itself. And this, I believe, is good news for philosophers of science.

Consequently, it is to that that we now turn, in the hope that philosophers will be who shall help it mature into a full-blown computational epistemology.

5 A LOGICAL TOOL OF AI

Classical logic has been developed to study well-defined, consistent, and unchanging mathematical objects. It thereby acquired a static character. *AI* needs to deal with knowledge in flux, and less than perfect conditions, by means of more dynamic forms of logic. Much of this has been the focus of research in logic programming, a field of *AI* which uses logic directly as a programming language [2], and provides specific implementation methods and efficient working systems to do so [3]. logic programming is moreover much used as a staple implementation vehicle for other *AI* approaches to logic.

What are the issues and requirements of the dynamics of logic which *AI*, namely via logic programming, has contributed to solve? [4]

[2] Cf. [Kowalski, 1979] for the foundations and [Alferes and Pereira, 1996] for a recent state-of-the art.

[3] For the most advanced, incorporating more recent theoretical developments, see the XSB system at: //www.cs.sunysb.edu/~sbprolog/xsb-page.html/.

[4] For a detailed technical treatment of the ensuing issues cf. our book [Alferes and Pereira, 1996], and references therein, and the more recent pointers in this text.

5.1 *Logic Programming*

There is a close connection between logical implication and physical causality. After all, logic permits us to model the goings on in the world. This is even more so in the case of Logic Programming, where all true conclusions must be supported, "caused", by some premises, and where implication is unidirectional, i.e. not contrapositive: "causes" do not run backwards. So-called Horn clause notation is used to express this directionality. This ability for logic programming to model actions by means of updates is brought out in the section on updating.

To produce a result or conclusion C we need a conjunction of enough positive conditions P to sustain it, conjoined by ',' to the absence or negation of a conjunction $\neg N$ of all the negative conditions that would prevent it, given P:

$$C \leftarrow P, \neg N$$

If there is more than one way to obtain C, then we have several rules of this form:

$$C \leftarrow P_i, \neg N_i$$

If our information about each rule, and about the whole set of rules for C, is complete, then we would have:

$$C \longleftrightarrow \bigvee_i P_i, \neg N_i$$

But what if we don't? First, we might not have enough information about each $\neg N_i$. We might conceivably know about each required P_i, without at least one of which C can never be achieved. But regarding the $\neg N_i$, i.e. all the conditions which, if present, would prevent us from concluding C in the presence of each P_i, that's more difficult.

5.2 *Open worlds*

Too many things can go wrong in an open non-mathematical world, some of which we don't even suspect. For all we know, some bomb might go off that destroys the whole physical setup we're trying to model logically (at a safe distance). Must we model that too? In the real world, any setting is too complex already for us to define it exhaustively each time. We have to allow for unforeseen exceptions to occur, based on new incoming information. Thus, instead of having to make sure or prove that some condition N_i is not present, we can assume it is not, on condition that we are prepared to accept subsequent information to the contrary, i.e. we may assume a more general rule than warranted, but must henceforth be prepared to deal with arising exceptions.

Take for example, with the obvious reading, this piece of knowledge, again in Horn clause form:

$$faithful(H,K) \leftarrow married(H,K), \neg lover(H,L)$$

We don't normally have explicit information about who is or is not the lover of whom, though that kind of information may arrive unexpectedly, especially of presidents[5]. Thus, we can code our knowledge on presumption that if someone H is married to K, and there is no information or basis to conclude $lover(H,L)$, then H is faithful to K:

$$faithful(H,K) \leftarrow married(H,K), not\ lover(H,L)$$

This is expressible via so-called default default negation $not\ P$, which may be read as "P is not provable". That is, we no longer require proof that $\neg lover(H,L)$ for any L, given some H, but will assume so unless we can establish $lover(H,L)$ for some L given H. In other words, if I have no evidence to conclude $lover(H,L)$ for some L given H, I can assume it false for all L given H.

5.3 Defeasible Assumptions

Default negation was introduced by *AI* researchers via Logic Programming, and may be read, interchangeably, as "P is not provable", or "the falsity of P is assumable", or "the falsity of P is abducible", or "there is no evidence for P", or "there is no argument for P". Default negation allows us to deal with lack of information, a common situation in the real world. It introduces non-monotonicity into knowledge representation. If later we're informed about a pair of lovers, we must then go back on our previous conclusion about whatever faithfulness we had presumed. Indeed, conclusions might not be solid because the rules leading to them may be. Legal texts, regulations, and courts employ this form of negation abundantly, as they perforce deal with open worlds, though it had not before been formalized in logic till *AI* too discovered its need of it.

5.4 Closed World Assumption

Mark that not should grant positive and negative information equal standing. That is, we should be able to write:

$$\neg faithful(H,K) \leftarrow married(H,K), not\ \neg lover(H,L)$$

to model instead a world where people are unfaithful by default or custom, and where it is required to explicitly prove that someone does not take any lover before we conclude that person not unfaithful. The issue arises because often, particularly in data bases and knowledge bases, the (CWA) is enforced. That is, the CWA

[5]At the time of writing this example I was in the USA, and the American president was being insistently harassed about a purported love affair.

obtains when we presume the database or knowledge base to have all the pertinent positive information, so that only what is explicitly stated, or explicitly derivable from it, is true; otherwise, it is presumed false. But how could it, in general, have all the pertinent information in a changing world?

Information is normally expressed positively, by dint of mental and linguistic economics. So, by CWA, the absent, non-explicitly obtainable information is usually the negation of positive information. Which means, when no information is available about lovers, that $\neg lover(H, L)$ is true by CWA, whereas $lover(H, L)$ is not. This asymmetry is undesirable, inasmuch as predicate names are purely conventional and we could, justifiably, have come up with positive predicate $non-lover(H, L)$ and its negation $\neg non-lover(H, L)$ to model our knowledge. Though in general P is considered the positive predicate if its extension greater than its negative counterpart, we might not know beforehand which is the greater one. A logical language should be indifferent with respect to such knowledge representation accidents.

5.5 *Explicit Negation*

In addition, when we have no factual or derivable information, either positive or negative, we'd like to be able to say that both are false epistemically, i.e. from the "knowledge we possess" point of view. Accordingly, the excluded middle postulate is unacceptable because some predication or other and its negation may be false simultaneously.

The CWA and the symmetry and epistemologic requisites, can be reconciled by reading '$\neg$' above not as classical negation, which complies with the excluded middle postulate stating that any predication is either true or false, but as yet a new form of negation, dubbed in logic programming ''explicit negation" [Alferes *et al.*, 1998a] (which ignores the excluded middle provision). Now we can state the CWA for just those predicates P or $\neg Q$ we wish, simply by writing:

$$\neg P \leftarrow not\ P \ \text{ or } \ Q \leftarrow not\ \neg Q$$

Alternatively, the use of a $not\ P$ or of a $not\ \neg Q$ assumption may be made at just those predicate occurrences requiring them, as we did before.

5.6 *Revising Assumptions*

Let us next examine the need for revising assumptions and for introducing a third truth-value, call it "undefined", into our framework.

When we combine the viewpoints of the two above worlds we become confused:

$$faithful(H, K) \leftarrow married(H, K), not\ lover(H, L)$$

$$\neg faithful(H, K) \leftarrow married(H, K), lover(H, L)$$

Assuming $married(H,K)$ for some H and K, it now appears that both $faithful(H,K)$ and $\neg faithful(H,K)$ are contradictorily true. Because we have no evidence for $lover(H,L)$ nor $\neg lover(H,L)$, there simply is no information about them, we make two assumptions about their falsity. But when an assumption leads to contradicition one should retract it. Yes, it is the venerable principle of *reductio ad absurdum, or* "reasoning by contradicition". In our case, two assumptions led to the contradicition. Which shall we retract? They are on equal footing!

5.7 Undefinedness

Given no other eventually preferential information, we retract both because we cannot justly decide between them. That is, we assume neither $lover(H,L)$ nor $\neg lover(H,L)$ false. Since neither is provably true also, we make each *undefined*, i.e. we introduce a third truth-value to better characterize this lack of information about some lovers' situation, thereby making $faithful(H,K)$ and $\neg faithful(H,K)$ undefined too. This imposition of undefinedness can be achieved simply, by adding to our knowledge:

$$\neg lover(H,L) \leftarrow \mathit{not}\ lover(H,L)$$

$$lover(H,L) \leftarrow \mathit{not}\ \neg lover(H,L)$$

Given no other information, we cannot prove either of $lover(H,L)$ or $\neg lover(H,L)$ true, or false. Any attempt to do so runs into a self-referential circle involving default negation. However, once we do hypothesize the one true the other perforce becomes false, and vice-versa. These two possible situations are not thus ruled out. But the safest, skeptical, third option is to take no side in this marital dispute, and abstain from believing either.

Indeed, the well-founded semantics of logic programs (WFS) assigns to the literals in the above two clauses the truth value *undefined* in its knowledge skeptical well-founded model, and allows also for the other two, non truth-minimal, more credulous models.

But why can we not simply add:

$$lover(H,L) \ \vee \neg lover(H,L) \ \ ?$$

Because it will not do. We would still not be able to prove either $lover(H,L)$ or $\neg lover(H,L)$ definitely true, and so the contradicition would ensue all the same! [6]

When dealing with non-provability one really needs a third truth-value to express our epistemic inability to come up with information. Even assuming the world is ontologically 2-valued, our access to information coded about it might be

[6] For a survey on paraconsistent *LP* semantics cf. [Damásio and Pereira, 1998].

3-valued in general. Besides, the world may very well not be 2-valued, or, in any case, not capturable in a 2-valued way. Can we say of some particle/wave that it is either here or not here? Is it really either here or not here, though we cannot say it? The inherent dispute apparently cannot be settled experimentally.

In any case we are in need of a third logical value for other reasons. As we build up our real world imperfect knowledge base, we may very well create, unwittingly and unawares, circular dependencies as above. For example, the Legislator may well enact conflicting, circular, laws. Still, we want to be able to carry on reasoning, whether or not such circularities legitimally express what they model. And when they do not, we want to detect and function with them rather than throwing away the baby with the bath water.

But undefinedness may also be legitimally pressed into service to express common knowlege, in its variously credulous and skeptical readings. Why, even yesterday I read in a book [Barth, 1996] of the old sailor proverb that ''He weathers the storms he cannot avoid, and avoids the storms he cannot weather", i.e.:

$$weather \leftarrow storm,\ not\ avoid$$

$$avoid \leftarrow storm,\ not\ weather$$

What will happen to an experienced sailor in a storm? Will he weather it or avoid it? Three outcomes are possible, including being undecided about the outcome. For us, it means we must weather circular knowledge, not avoid it in the calm waters of the artificial mathematical paradises...

5.8 *Expressiveness*

The introduction of this 3-valued semantics into logic programming is an important innovation of *AI*, with a number of consequences. Namely, as we have seen, it allows us to be skeptical and suppose no more than is warranted.

Another innovation, to recapitulate, is that the negation $\neg$ used above, called explicit negation, is 3-valued too, and so differs from classical negation. Explicit negation, we've seen, does not conform to the principle of excluded middle. Predications and their negations can be both false ("the chair is angry", "the chair is not angry"). In a knowledge base, we may definitely not be able to prove either something or its (explicit) negation, so that both become false by default.

Additionally, the introduction of explicit negation, and in particular its combination with default negation, provides for more expressive knowledge representation. Consider for instance the injunction

$$\neg cross \leftarrow train$$

versus the alternative injunction

$$\neg cross \leftarrow not\ \neg train$$

The latter is clearly a safer option. According to it you must actually prove that a train is not coming before $\neg cross$ becomes false, whereas with the first option $\neg cross$ becomes false solely on the grounds that you fail to prove a train is coming!

Yet another innovation is that the implicational arrow $\leftarrow$, when combined with default or explicit negation, is not material implication. Contrapositives are not countenanced. Consider:

$$\begin{array}{lll} A & \leftarrow & B \\ & & \neg A \end{array}$$

$\neg B$ does not follow, for there is no presumption of the contrapositive $\neg B \leftarrow \neg A$. Whenever contrapositives are desired they must be explicitly added. Instead, $\leftarrow$ should be seen as expressing an inference rule, which can be used procedurally "bottom-up" to conclude A given B, or "top-down", like an invoked procedure, to try and prove the body B in order to prove the head A. Facts are just procedures with empty body. This inferential reading of clauses as rules turns $\leftarrow$ into a directional operator, which makes it rather appropriate to model not only inference but causality as well.

Indeed, the semantics of logic programs emphasizes exactly this, with their insistence on admitting only minimal models. Minimal models are those for which every positive or explicitly negated literal *true* in the model is supported on some rule whose head or conclusion is the literal, and whose body or set of premises is in turn supported. Thus facts, since they have an empy body, are automatically supported.

Additionally, any positive or any explicitly negated literal is *false* just in case all rules for it have a false body. Consequently, if there are no rules whatsoever for a literal it is automatically false.

All other positive or explicitly negated literals are *undefined* (this can happen only if they are involved in unresolved self-referential loops through default negation). Finally, a default negated literal, $not\ P$, is *true/ false/ undefined* just in case P is, respectively, *false/ true/ undefined*.

5.9 Theory Diagnosis

Many examples exist of the use of default negation, explicit negation, contradicition removal, and 3-valuedness, made possible by research in Logic Programming, namely in the areas of abduction, argumentation, belief revision, knowledge updating, learning, and diagnosis [Alferes and Pereira, 1996; Gartner *et al.*, 2000]. Let us illustrate the latter, i.e. the diagnosis or debugging of a generic knowledge base. Its connections with the other aforementioned areas allow us to touch on them too.

Suppose we have the following rules, expressing a logic program that allows us to compute predicate C on the basis of predicates P and N:

$$C(X) \leftarrow P(X), \neg N(X)$$

$$\begin{array}{ll} P(a) & \neg N(a) \\ P(b) & \\ P(c) & \neg N(c) \end{array}$$

Let us allow for the chance that the rule for C is conceivably wrong. i.e. that it may have exceptions, by writing it thus:

$$C(X) \leftarrow P(X), \neg N(X),\ not\ exception(C(X))$$

If nothing else is stated about the $exception$ predicate, the conclusions of the logic program remain unchanged, so the program may contain these rules from the start, for any predicate we wish.

If we next learn that $C(a)$ is false, we can salvage the rule for C, since it is no longer valid for all $X = a$, simply by adding to our knowledge:

$$exception(C(a))$$

The rule for C will continue to work for the remaining cases, though more exceptions might subsequently be accumulated.

This method takes care of rules that are unsound, i.e. produce wrong results.

What about the other problematic case, that of incompleteness, i.e. when results that would be correct are missing? Suppose we learn that $C(b)$ should true, but that our rules rule do not account for it? Well, it is enough to introduce for every set of rules for a predicate a catch-all rule. In this case:

$$\begin{array}{rcl} C(X) & \leftarrow & missing(C(X)) \\ P(X) & \leftarrow & missing(P(X)) \\ \neg N(X) & \leftarrow & missing(\neg N(X)) \end{array}$$

If nothing else is stated about the $missing$ predicate, the conclusions of the logic program remain unchanged, so the program may contain these rules from the start. To make $C(b)$ true, it now suffices to abduce, or adopt, one of two hypotheses: that $missing(C(b))$ or that $missing(\neg N(b))$ is true.

This method takes care of rules that are incomplete, i.e. fail to produce results.

We have achieved generality. By making the two predicates, *exception*(_) and *missing*(_), abducible ones, we can diagnose and debug a knowledge base expressed in logic.

Abduction is a well-known reasoning process by which one can adopt hypotheses to the effect that some predicate instances (or their negations) are true in order to prove some conclusion. It's reasoning from goals to requirements, and has been

extensively studied in the logic programming context[7]. Of course, the assumptions or hypotheses so adopted may later prove erroneous, if and when they lead to contradicition. We have then to make provision for the application of a contradicition removal process based on revising assumptions, and possibly adopting some other assumptions instead. This too has been extensively studied in *AI* and in the logic programming setting, and as a result automated reasoning systems have been made available which do that work for us [Alferes *et al.*, 1995; Damásio *et al.*, 1995] and have been applied to real domains [Froehlich *et al.*, 1999].

5.10 Updating

Last but not least, work in logic programming has concerned itself with the updating of a knowledge base by another one. This notion of knowledge updating, as opposed to that of simple fact updating, opens up another dimension to the dynamics of logic, in contradistinction to the statics of the logic of old. Given an existing knowledge base, containing facts and rules, and some new knowledge with facts and rules as well, possibly contradicting the previous knowledge when added to it, what is the resulting updated knowledge base? Can this process be iterated?

Most of the work conducted so far in the field of logic programming has focused on representing *static* knowledge, i.e., knowledge that does not evolve with time. This is a serious drawback when dealing with *dynamic knowledge bases* in which not only the *extensional* part (the set of facts) changes dynamically but so does the *intensional* part (the set of rules). Recent work has shown how this can be achieved with generality [Alferes *et al.*, 1998; Alferes and Pereira, 2001]. Application areas in point are when the two knowledge bases consist in pieces of legislation, or regulations, or safety procedures, or rules of robot conduct.

The introduction of dynamic program updates extends program updates to ordered sets of logic programs (or modules). When this order is interpreted as a time order, dynamic program updates describe the evolution of a logic program which undergoes a sequence of modifications. This opens up the possibility of incremental design and evolution of logic programs, leading to the paradigm of *dynamic logic programming* (*DLP*). This new paradigm significantly facilitates modularization of logic programming, and thus modularization of non-monotonic reasoning as a whole.

Specifically, suppose that one is given a set of logic program modules, each describing a different state of our knowledge of the world. Different states may represent different time points or different sets of priorities or perhaps even different viewpoints. The rôle of the dynamic program update is to employ the mutual relationships existing between different modules to precisely determine, at any given module composition stage, the declarative as well as the procedural semantics of the combined program resulting from the modules [Alferes *et al.*, 2000]. However,

[7] For an overview, cf. [Kakas *et al.*, 1993]. For an implementation, cf. [Alferes *et al.*, 1999a].

ordered sets of program modules need not necessarily be seen as just a temporal evolution of a logic program. Different modules can also represent different sets of priorities, or viewpoints of different agents. In the case of priorities, a dynamic program update specifies the exact meaning of the *"union"* of the modules, subject to the given priorities.

In order to represent negative information, *generalized logic programs* are employed in updating. These allow default negation not only in rule bodies but also in their heads. This is needed, in particular, in order to specify that some atoms should became false, i.e., should be deleted. However, such updates are far more expressive than a mere insertion and deletion of facts. The updates can be specified by means of arbitrary program rules, and thus the updates themselves are logic programs. Consequently, the approach empowers one to update one generalized logic program P (the initial program) by another generalized logic program U (the updating program), obtaining as a result a new, updated logic program $P \oplus U$. As a consequence, program modules may contain mutually contradictory as well as overlapping information.

The common intuition behind the update of a model of the world has been based on what is referred to as the commonsense law of inertia, i.e. things do not change unless they are expressly made to do so. Suppose, for example, that we have a model in which "sunshine" is true and "rain" is false; if later we receive the information that the sun is no longer shining, we conclude that "sunshine" is false, due to the update, and that "rain" is still false by inertia.

Suppose now that our vision of the world is described instead by a logic program and we want to update it. Is updating each of its models enough? Is all the information borne by a logic program contained within its set of models? The answer to both these questions is negative. A logic program encodes more than the set of its individual models. It encodes the relationships between the elements of a model, which are lost if we envisage updates simply on a model by model basis, as proposed by revision programs. In fact, while the semantics of the resulting knowledge base after an update indeed represents the intended meaning when just the extensional part of the original knowledge base (the set of facts) is being updated, it leads to strongly *counter-intuitive* results when also the intensional part of the database (the set of rules) undergoes change, as the following example shows.

Consider the logic program P :

$$sleep \leftarrow not\ tv\_on$$

$$tv\_on \leftarrow$$

$$watch\_tv \leftarrow tv\_on$$

Clearly $M = \{tv\_on, watch\_tv\}$ is its only stable model. Suppose now that the update U states that there is a power failure, and that if there is a power failure then the TV is no longer on, as represented by the logic program U:

$$not\ tv\_on \leftarrow power\_failure$$

$$power\_failure \leftarrow$$

With simple model updating, we would have $M_U = \{power\_failure, watch\_tv\}$ as the only update of M by U. This is because $power\_failure$ needs to be added to the model and its addition forces us to make tv_on false. As a result, even though there is a power failure we are still watching TV.

However, by inspecting the initial program and the updating rules, we are likely to conclude that since $watch\_tv$ was true only because tv_on was true, the removal of tv_on should make $watch\_tv$ false by default. Moreover, one would expect $sleep$ to become true as well. Consequently, the intended model of the update of P by U is the model $M_{U'} = \{power\_failure, sleep\}$. Suppose now that another update U_2 follows, described by the logic program:

$$not\ power\_failure \leftarrow$$

stating that the power is back up again. We should now expect the TV to be on again. Since power was restored, i.e. $power\_failure$ is false, the rule $not\ tv\_on \leftarrow power\_failure$ of U should have no effect, and the truth value of tv_on should be obtained by inertia from the rule $tv\_on \leftarrow$ of the original program P.

This example illustrates that, when updating knowledge bases, it is not sufficient to just consider the truth values of literals figuring in the heads of its rules, because the truth value of their rule bodies may also be affected by the updates of other literals. In other words, it suggests that the *principle of inertia* should be applied not just to the individual literals in an interpretation but rather to the *entire rules of the knowledge base*.

Newton's first law, also known as the law of inertia, states that: *"every body remains at rest or moves with constant velocity in a straight line, unless it is compelled to change that state by an unbalanced force acting upon it"* (adapted from [Newton, 1726]). One often tends to interpret this law in a commonsensical way, as things keeping as they are unless some kind of force is applied to them. This is true but it doesn't exhaust the meaning of the law. It is the result of all applied forces that governs the outcome. Take a body to which several forces are applied, and which is in a state of equilibrium due to those forces canceling out. Later, one of those forces is removed and the body starts to move.

The same kind of behaviour presents itself when updating programs. Before obtaining the truth value, by inertia, of those elements not directly affected by the update program, one should verify whether the truth of such elements is not indirectly affected by the updating of other elements. That is, the body of a rule may act as a force that is sustaining the truth of its head, but this force may be withdrawn if and when the body becomes false. Accordingly, the correct way to view program updating, and in particular the rôle of inertia, is to say that it is the rules of the initial program that carry over to the updated program, due to inertia (instead of the truth of literals) just in case they are not overruled by the updating program. This should be so because rules encode more information than just their models.

This approach was first adopted in [Leite and Pereira, 1997], where the authors presented a program transformation which, given an initial program and an update program, produces an updated program obeying rule inertia. The stance of the dynamic logic programming approach is precisely that of ensuring that any added update rules are indeed in force, and that previous program rules are still in force (by inertia) only in so far as possible, i.e. they are kept for as long as they do not conflict with any newly added ones.

To represent negative information in logic programs and their updates *DLP* allows for the presence of default negation in rule heads. It is worth noting why, in the update setting, generalizing the language to allow default negation in rule heads is more adequate than introducing explicit negation in programs (both in heads and bodies). Suppose we are given a rule stating that A is true whenever some condition $Cond$ is met. This is naturally represented by the rule $A \leftarrow Cond$. Now suppose we want to say, as an update, that A should no longer be the case (i.e. should be deleted or retracted), if some condition $Cond'$ is met. How to represent this new knowledge?

By using extended logic programming (with explicit negation) this could be indicated by $\neg A \leftarrow Cond'$. But this rule says more than we want to. It states that A is false upon $Cond'$, and we just want to go as far as to say that the truth of A is only to be deleted in such a case. All is wont said is that if $Cond'$ is true then $not\ A$ should be the case, i.e. $not\ A \leftarrow Cond'$. The difference between explicit and default negation is fundamental whenever the information about some atom A cannot be assumed to be complete, as we have seen before. Under these circumstances, the former means that there is evidence for A being false, while the latter means that there is no evidence for A being true. In the deletion example, we desire the latter case.

In other words, a $not\ A$ head means A is deleted if the body holds. Deleting A means that A is no longer true, not necessarily that it is false. When the CWA is adopted as well, then this deletion causes A to be false. In the updates setting, the CWA must be explicitly encoded from the start, by making all $not\ A$ literals false in the initial program being updated. That is, the two concepts, deletion and CWA, are orthogonal and must be separately incorporated. In the *Stable Models* and *Well-Founded* semantics of single generalized programs, the CWA is adopted *ab initio*, and default negation in the heads is conflated with non-provability because there is no updating, and thus no deletion. Note that in updates the head $nots$ cannot be moved freely into the body, to obtain simple denials, for two reasons: first, this manoeuvre, although allowed in a 2-valued semantics does not carry over to a 3-valued one; second and more importantly, there is inescapable pragmatic information in specifying exactly which not literal figures in the head, namely the one being deleted when the body holds true, and thus it is not indifferent that any other (positive) body literal in the denial could be moved to the head.

In [Alferes *et al.*, 1999], a Language of Updates (*LUPS)* has been proposed, expressly designed for specifying sequences of updates to logic programs. Even

though the main motivation behind the introduction of *DLP* was to represent the evolution of knowledge in time, the relationship between the different states can encode other aspects of a system of knowledge, as pointed out above. In fact, since their introduction, *DLP* and *LUPS* have been employed to represent several distinct aspects, namely as a way to:

- represent and reason about the evolution of knowledge in time [Alferes *et al.*, 2000; Alferes *et al.*, 2000b];
- combine rules learned by agents [Lamma *et al.*, 2000a];
- reason about updates of agents' beliefs [Dell'Acqua and Pereira, 1999];
- model agent interaction [Quaresma and Pereira, 1998; Quaresma and Rodrigues, 1999];
- model and reason about actions [Alferes *et al.*, 2000a; Alferes *et al.*, 2000b];
- resolve inconsistencies in metaphorical reasoning [Leite *et al.*, 2000];
- combine updates and preferences [Alferes and Pereira, 2000a].

The common aspect among these applications of *DLP* is that the states associated with the given set of theories only encode one of the several possible representational dimensions (e.g. time, hierarchies, domains,...). This is so because *DLP* is only defined for linear sequences of states. For example, *DLP* can be used to model the relationship of a hierarchically related group of agents, and *DLP* can be used to model the evolution of a single agent over time. But *DLP*, as it stands, cannot deal with both settings at once, and model the evolution of such a group of agents over time.

An instance of such a multi-dimensional scenario can be found in legal reasoning, where the legislative agency is divided conforming to a hierarchy of power, governed by the principle *Lex Superior* (Lex Superior Derogat Legi Inferiori) according to which the rule issued by a higher hierarchical authority overrides the one issued by a lower one, and the evolution of law in time is governed by the principle *Lex Posterior* (Lex Posterior Derogat Legi Priori) according to which the rule enacted at a later point in time overrides the earlier one.

In effect, knowledge updating is not to be simply construed as taking place in the time dimension alone. Several updating dimensions may combine simultaneously, with or without the temporal one, such as specificity (as in taxonomies), strength of the updating instance (as in the legislative domain), hierarchical position of knowledge source (as in organizations), credibility of the source (as in uncertain, mined, or learnt knowledge), or opinion precedence (as in a society of agents). For this to be possible, *DLP* has been extended to allow a more general structure of states.

Multi-dimensional Dynamic logic programming (*MDLP*) [Leite *et al.*, 2001a] generalizes *DLP* to allow for collections of states represented by arbitrary acyclic

digraphs, not just sequences of states. *MDLP* assigns semantics to sets and subsets of logic programs, depending on how they stand in relation to one another, as defined by the acyclic digraph (*DAG*) that represents the states and their configuration. By dint of such natural generalization, *MDLP* affords extra expressiveness, thereby enlarging the latitude of logic programming applications unifiable under a single framework. The generality and flexibility provided by a *DAG* ensures a wide scope and variety of new possibilities. By virtue of the newly added characteristics of multiplicity and composition, *MDLP* provides a “societal” viewpoint in logic programming, important in these web and agent days, for combining knowledge in general.

Updating inevitably raises issues about revising and preferring, and some work is emerging on the articulation of these distinct but highly complementary aspects. And learning is usefully seen as successive approximate change, as opposed to exact change, and combining the results of learning by multiple agents, multiple strategies, or multiple data sets, inevitably poses problems within the province of updating. Finally, goal directed belief revision can be fruitfully construed as abductive updating.

Thus, not only do the aforementioned topics combine naturally together — and so require precise, formal, means and tools to do so — but their combination results in turn in a nascent complex architectural basis and component for logic programming rational agents, which can update one another and common, structured, updatable blackboard agents. It can be surmised, consequently, that the fostering of this meshing of topics within the Logic Programming community is all of opportune, seeding, and fruitful. Indeed, application areas such as software development, multi-strategy learning, abductive belief revision, model-based diagnosis, agent architecture, and others, are being successfully pursued while employing precisely this outlook.

6 A DYNAMICAL LOGIC FRAMEWORK

It is not too difficult to imagine how a combined process of rule generation, of systematic diagnosis, and of rule revision by updating, can be used to achieve automated theory learning, in an integrated way, within the uniform setting of logic programming.

To initiate the learning, one starts with some fixed, already acquired, background knowledge in rule form, i.e. a theory, and with a rule generator to add to it new purported knowledge, in order to explain abductively known observations, whether positive or negative, in the form of facts and explicitly negated facts.

The goal is to generate rules that define a positive concept as well as its negated concept, so that they cover all known observation instances. This automatic generation of new rules is subjected to a pre-defined bias, i.e. only some rule forms, and predicates comprising them, are allowed in the generation process. Newly generated rules may contradict one another, on some of the observation instances, and

so they must be subjected to a diagnosis, to identify alternative possible minimal revisions.

To decide which revision to adopt next, desirable new possible observations are conceived of, with respect to the ongoing available theory, and whose results, if known, would allow us to decide among the competing revisions. The results of these so-called crucial observations are then obtained, either by the program soliciting them, or by a belief revision process subsequently executed to carry out the actions leading to the observation results

Once the desired revisions are selected on the basis of the results, and of also programmed preference criteria, the revisions are enacted by an update procedure. Note that revised rules can themselves be subjected to later revisions if needed.

Indeed, the whole process will be iterated on the basis of new incoming knowledge, or by knowledge confrontation of among differently evolved automated theories, with distinct backgrounds, biases, rule generators, diagnosers, revisors, preferences, planners, observations, and updating procedures, comprising a rational agent.

Agent epistemic confrontation relies on argumentation and mutual debugging. Besides the legislative and legal domain, the province of scientific discussion, too, relies on such procedures, and can benefit from their automation.

Tackling argumentation involves a set of tools similar to those of diagnosis and debugging. Arguments can attack another argument's assumptions either directly, by proving the negation of an assumption, or indirectly, by contradicting a conclusion of other argument which rests on its assumptions. However, such attacks by an argument upon another can, in turn, be counter-attacked in the same way, and may in response counter-counter-attack, etc. Logic programming has shown how this process can be studied and conclusions drawn about competing mutually contradictory arguments, and how they each can be revised to reach agreement.

The above sketch of a dynamical logical framework is being made concrete within the context of our *MENTAL* agents project[8], at the Centro de Inteligência Artificial, Universidade Nova de Lisboa.

In both academia and industry it is increasingly felt that intelligent agents will be a key technology as computing systems become ever more distributed, interconnected, and open. In such environments, the ability of agents to autonomously plan and pursue their actions and goals, to cooperate, coordinate, and negotiate with others, and to respond flexibly and intelligently to dynamic and unpredictable situations will lead to significant improvements in the quality and sophistication of the software systems that can be conceived and implemented, and the application areas and problems which can then be addressed.

The aim of the *MENTAL* project is that of establishing, on a sound theoretical basis, the design of an overall architecture for mental agents (i.e. comprising

[8]Portuguese research funding programme *PRAXIS XXI*, project *MENTAL* (2/2.1/TIT/1593/95), http://centria.di.fct.unl.pt/˜ica/mental/.

knowledge, beliefs, and intentions) based on, and building on the strengths of logic programming.

The agents share a common changeable environment, itself modeled in logic programming. The whole framework is being conceptualized in logic programming, and experimented with by means of distributed logic programming [Leite *et al.*, 2001]. The use of agents allows for information to be distributed, and shared, only as the need arises. Thus knowledge manipulation complexity can be reduced. Efficiency can be gained too by executing agents concurrently.

An agent must be able to manage its knowledge, T beliefs, intentions (goals), and plans, as it receives new information and instructions, and to react to changing conditions in the environment. It must also be capable of interacting with other agents, by exchanging knowledge and beliefs, as well as of reacting to other agents' requests. Any two such agents will be able to cooperate either by diagnosing errors or missing information in each others' information base, or to cooperate in the use of common resources, in avoiding undesired mutual interference in their plans, and in joint belief revision to achieve a common goal.

Every agent is composed of specialized, possibly concurrent, function related subagents, that execute the various goals and instructions assigned to it by the agent. Examples of such subagents are those implementing the reactive, reasoning, belief revision, argumentation, explanation, learning, dialogue management, information gathering, preference evaluation, strategy, and diagnosis functionalities. The subagents are coordinated by a meta-level layer that ensures internal task distribution and belief revision, subagent communication, final decisions, and all interaction with the outside, i.e., with both the environment and other agents, including communication, interrupt handling, requests, and observations.

Although each agent has a meta-level coordination, a collection of interacting agents has no such abode, and their collective behaviour will manifest (emergent) properties that are difficult to foresee.

Because knowledge and belief are in general incomplete, contradictory, and error prone — and all the more so in the multi-agent setting — we make use of semantics, procedures, and implementations, for dealing with contradictory, incomplete, erroneous, imperfect, vague and default information, abduction, belief revision, debugging, and argumentation.

We attain reinforcement learning by using belief revision techniques for achieving an effect similar to back propagation. Such learning allows the evolving of the strength of evidence attached to an agent's knowledge and beliefs, as a result of their contribution to correct or incorrect conclusions. Furthermore, agents may compare and combine their degrees of evidence, either to argue, to partake of information, or to reach consensus. The use of genetic algorithms for evolving belief "genes" (or "memes") is also a technique we have explored for learning, and it opens up a new emergent field: that of linking the genetic algorithm approach to

logic based knowledge evolution in an agent. Moreover, the exchange of genetic material enables agents to cross-fertilize experiences.[9]

The use of logic programming for the overall endeavour is justified on the basis of it providing a rigorous single encompassing theoretical basis for the aforesaid topics, as well as an implementation vehicle for parallel and distributed processing. Additionally, logic programming provides a formal high level flexible instrument for the rigorous specification and experimentation with computational designs, making it extremely useful for prototyping, even when other, possibly lower level, target implementation languages are envisaged.

7 CONCLUDING REMARKS

How can the general reasoning issues explored here be approached, within a uniform framework, except in the context of logic? I submit indeed that logic is the only conceivable venue for the core of such an enterprise. And must we not do it? Must we not mechanize reasoning procedures so that robots and computers can perform, for us or with us, the sometimes tedious, sometimes overly complex work involved in them? Surely, only so persisting will permit the grappling of the evermore demanding reasoning problems that face us.

I hope to have convinced you that *AI*, most especially through logic programming, will continue to accomplish a good deal in identifying, formalizing, and implementing the Laws of Thought. Most notably, *AI* has taken on the challenge of opening up logic to the dynamics of knowledge in flux. And in so doing, it has been progressively meeting our expections and requirements. Continuing this work is essential if we are to cope, albeit with the aid of computers, with the challenges of evermore accumulated and distributed knowledge, in a changing world.

A wealth of investment in logic programming research has accumulated in the form of published results and built systems, over the more than 25 years since its inception. Some of these could not be brought to full fruition at the time because of technological costs and funding limitations, or more simply because of the lack of cheap and efficient requisite equipment, nowadays having become affordable and with better performance. So, though the theory and implementational know-how were there at the time, their full fruition was then not viable.

However, to-day, younger researchers are not well aware of such past results and of their potential, some from the early eighties, even when the implementational technology may have evolved to allow them to be put to use[10]. And all the more so because the complexity of present systems makes such immediately available "theory technologies" all the more desirable. On the other hand, the cost of memory and the appearance on the scene of new "implementation technologies", such

[9] This new approach is reported in [Lamma *et al.*, 2000; Lamma *et al.*, 2000a].

[10] Illustrative case examples in point are those of "intelligent backtracking" and of "declarative debugging".

as (theory-driven) tabling, make the prior investment in theory not only realizable now but affordable and profitable as well.

Conversely, we should not stop or refrain from investing now in the future, for the "theory technologies" of today, in addressing real felt problems, will be the providers of tomorrow's implementational and applicational reapers.

For instance, were it not for the impressive development in Logic Programming semantics in the past 12 years or so, and its extensions to non-monotonic and other forms of reasoning, we would not have today the impressive, sound, and efficient system implementations of such semantics (be it the stable or well-founded varieties or their hybrid combination), which have been opening up a whole new gamut of application areas as well as a spate of sophisticated reasoning abilities over them. This was made possible only through successive and prolonged efforts at theoretical generalization and synthesis, and by way of the combined integration of theory, procedure development, and practical implementation.

The lesson, then, for further investing in this still emerging information technology, is that it must be continual: both for the fruition of past investment as for the seeding of timely future delectation.

The logic programming paradigm provides a well-defined, general, integrative, encompassing, and rigorous framework for systematically studying computation, be it syntax, semantics, procedures, or attending implementations, environments, tools, and standards. *LP* approaches problems, and provides solutions, at a sufficient level of abstraction so that they generalize from problem domain to problem domain. This is afforded by the nature of its very foundation in logic, both in substance and method, and constitutes one of its major assets.

Indeed, computational reasoning abilities such as assuming by default, abducing, revising beliefs, removing contradicitions, updating, belief revision, learning, constraint handling, etc., by dint of their generality and abstract characterization, once developed can readily be adopted by, and integrated into, distinct topical application areas.

Logic programming is, without a doubt, the privileged melting pot for the articulate integration of functionalities and techniques, pertaining to the design and mechanization of complex systems, addressing ever more demanding and sophisticated computational abilities. For instance, consider again those reasoning abilities mentioned in previous sections. Forthcoming rational agents, to be realistic, will require an admixture of any number of them to carry out their tasks. No other computational paradigm affords us with the wherewithal for their coherent conceptual integration. And, all the while, the very vehicle that enables testing its specification, when not outright its very implementation. Or is there?

Moreover, the issue is not just that of conceptual integration but that of conceptual cross-fertilization too. How can one employ learning in the service of belief revision? How does one remove contradictions from updates? How can learnt preferences be combined with preference updates to guide belief revision in a rational agent? How may fuzzy preferences be revised in the light of ongoing

abduced assumptions and constraints? How can probabilistic and multi-valued logic programs be combined?[11]

Participation by philosophers in this venture can take the form of partaking in the development of the conceptual underpinnings of the required computer implemented logic tools, and of employing such tools for the computer functional rational reconstruction of historical epistemological argument[12], and of scientific theory evolution duplication, by automated learning. Evolving taxonomic classification of natural kinds might be a good starting point.

8 ACKNOWLEDGEMENTS

This chapter is a much extended and updated version of the paper titled "The Logical Impingement of AI", first published in [Pereira, 1999].

This work was mostly supported by PRAXIS XXI project MENTAL (2/2.1 /TIT/1593/95), and by NATO scholarships while the author was on sabbatical leaves at the Department of Computer Science, University of California, Riverside. Abundant thanks are due to the several co-authors of joint work mentioned.

BIBLIOGRAPHY

[Alferes and Pereira, 1996] José Júlio Alferes and Luís Moniz Pereira. *Reasoning with Logic Programming*, Lecture Notes in Artificial Intelligence, vol. 1111, Springer, 1996.

[Alferes and Pereira, 2000a] José Júlio Alferes and Luís Moniz Pereira. Updates plus preferences. In M. O. Aciego, I. P. de Guzmán, G. Brewka and L. M. Pereira, eds. *Logics in AI, Procs. JELIA'00*, Málaga, LNAI 1919, pages 345–360. Springer, 2000.

[Alferes and Pereira, 2001] José Júlio Alferes and Luís Moniz Pereira. Logic programming updating: a guided tour. In *Essays in Honour of Robert Kowalski*, Fariba Sadri and Antonis Kakas (eds.), Springer, 2001.

[Alferes *et al.*, 1995] José Júlio Alferes, Carlos V. Damásio and Luís Moniz Pereira. A logic programming system for non-monotonic reasoning. *Journal of Automated Reasoning*, 14:93-147, 1995.

[Alferes *et al.*, 1998] José Júlio Alferes, João A. Leite, Luís Moniz Pereira, Halina Przymusinska and Teodor C. Przymusinski. Dynamic logic programming, *Int. Conf. on Knowledge Representation and Reasoning (KR'98)*, Trento, Italy. Stuart Shapiro (ed.), Morgan Kaufmann, 1998.

[Alferes *et al.*, 1998a] José Júlio Alferes, Luís Moniz Pereira and Teodor C. Przymusinski. 'Classical' negation in nonmonotonic reasoning and logic programming, *Journal of Automated Reasoning*, 20:107-142, 1998.

[Alferes *et al.*, 1999] José Júlio Alferes, Luís Moniz Pereira, H. Przymusinska and T. C. Przymusinski. LUPS - a language for updating logic programs. In M. Gelfond, N. Leone and G. Pfeifer (eds.), *Procs. of the 5th Int. Conf. on Logic Programming and Nonmonotonic Reasoning*, El Paso, Texas USA, pages 162-176, LNAI 1730, Springer, 1999. An extended updated version has been accepted by the AI journal as LUPS - a language for updating logic programs.

[Alferes *et al.*, 1999a] José Júlio Alferes, Luís Moniz Pereira and Terrance L. Swift. Well-founded Abduction via Tabled Dual Programs. In D. De Schreye (ed.), *Procs. 16th Int. Conf. on Logic Programming*, Las Cruces, New Mexico, pages 426-440, MIT Press, 1999.

[11]Cf. [Damásio and Pereira, 2000; Damásio and Pereira, 2001; Damásio and Pereira, 2001a; Damásio *et al.*, 1999].

[12]As can be found for instance in [Lakatos, 1976].

[Alferes *et al.*, 2000] José Júlio Alferes, João A. Leite, Luís Moniz Pereira, Halina Przymusinska and Teodor C. Przymusinski. Dynamic Updates of Non-Monotonic Knowledge Bases, *The Journal of Logic Programming*, 45(1-3): 43-70, September/October 2000.

[Alferes *et al.*, 2000a] José Júlio Alferes, João A. Leite, Luís Moniz Pereira, Paulo Quaresma. Planning as Abductive Updating. In D. Kitchin (ed.), *Procs. AISB'00 Symposium on AI Planning and Intelligent Agents*, pages 1-8, AISB, Birmingham, England, April 2000.

[Alferes *et al.*, 2000b] José Júlio Alferes, Luís Moniz Pereira, Halina Przymusinska, Teodor C. Przymusinski and Paulo Quaresma. Dynamic Knowledge Representation and its Applications. In Stefano Cerri and Danail Dochev (eds.), *Procs. of The 9th Int. Conf. on Artificial Intelligence - Methodology, Systems, Applications (AIMSA'00)*, Varna, Bulgaria, LNAI 1904, Springer, pages 1-10, 2000.

[Barth, 1996] John Barth. *On with the Story*, page 55, Back Bay Books, Little Brown and Company, 1996

[Damásio *et al.*, 1995] Carlos V. Damásio, Wolfgang Nejdl, Luís Moniz Pereira and Michael Schroeder. Model-based diagnosis preferences and strategies representation with logic metaprogramming. In K. Apt and F. Turini (eds.), *Meta-Logics and Logic Programming*, pp. 267-308, MIT Press, 1995.

[Damásio and Pereira, 1998] Carlos V. Damásio and Luís Moniz Pereira. A Survey on Paraconsistent Semantics for Extended Logic Programas. In D.M. Gabbay and Ph. Smets (eds.), *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, vol. 2, pp 241-320, Kluwer Academic Publishers, 1998.

[Damásio and Pereira, 2000] Carlos V. Damásio and Luís Moniz Pereira. Hybrid probabilistic logic programs as residuated logic programs. In M. O. Aciego, I. P. de Guzmán, G. Brewka and L. M. Pereira (eds.), *Logics in AI, Procs. JELIA'00*, Málaga, LNAI 1919, pages 57-72, Springer 2000.

[Damásio and Pereira, 2001] Carlos V. Damásio and Luís Moniz Pereira. Monotonic and residuated logic programs. In *Procs. 6th European Conf. on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU'01)*, S. Benferhat and Ph. Besnard, Springer, LNAI 2001.

[Damásio and Pereira, 2001a] Carlos V. Damásio and Luís Moniz Pereira. Antitonic logic programs. In *Procs. 6th Int. Conf. on Logic Programming and Nonmonotonic Reasoning (LPNMR'01)*, T. Eiter and M. Truszczynski (eds), Springer LNAI 2001.

[Damásio *et al.*, 1999] Carlos V. Damásio, Luís Moniz Pereira and Terrance L. Swift. Coherent well-founded annotated logic programs. In M. Gelfond, N. Leone and G. Pfeifer (eds.), *Procs. 5th Int. Conf. on Logic Programming and Nonmonotonic Reasoning (LPNMR'99)*, El Paso, Texas USA, pages 262-276, LNAI-1730, Springer, 1999.

[Deacon, 1997] Terrence Deacon. *The Symbolic Species*, W.W. Norton, 1997.

[Dell'Acqua and Pereira, 1999] Pierangelo Dell'Acqua, Luís Moniz Pereira. Updating agents. In S. Rochefort, F. Sadri and F. Toni (eds.), *Procs. ICLP'99 Workshop on Multi-Agent Systems in Logic (MASL'99)*, Las Cruces, New Mexico, 1999.

[Dell'Acqua and Pereira, 2001] Pierangelo Dell'Acqua, Luís Moniz Pereira. Preferring and updating in abductive multi-agent systems. In *Procs. Workshop on Engineering Societies in the Agent's World 2001 (ESAW'01)*, Andrea Omicini et al. (eds), Prague.

[Donald, 1991] Merlin Donald. *Origins of the Human Mind*, Harvard University Press, 1991.

[Froehlich *et al.*, 1999] Peter Froehlich, Carlos V. Damásio, Wolfgang Nejdl, Luís Moniz Pereira and Michael Schroeder. Using extended logic programming for alarm-correlation in cellular phone networks. In *Procs. of The 12th International Conference on Industrial & Engineering Applications of Artificial Intelligence & Expert Systems IEA/AIE-99*. Cairo, Egypt. LNAI, Springer 1999.

[Gartner *et al.*, 2000] Joseph Gartner, Terrance L. Swift, Allen Tien, Carlos V. Damásio and Luís Moniz Pereira. Psychiatric diagnosis from the viewpoint of computational logic. In J. Lloyd et al. (eds.), *Procs. of First Int. Conf. on Computational Logic (CL 2000)*, London, UK, pages 1362-1376, LNAI 1861, Springer, July 2000.

[Gross and Levitt, 1994] Paul R. Gross and Norman Levitt. *Higher Superstition*, The Johns Hopkins University Press, 1994.

[Holland, 1998] John Holland. *Emergence*, Addison-Wesley, 1998.

[Kakas *et al.*, 1993] Anthony C. Kakas, Robert A. Kowalski and Francesca Toni. Abductive logic programming, *Journal of Logic and Computation*, 2:719-770, 1993.

[Kowalski, 1979] Robert A. Kowalski. *Logic for Problem Solving*, North-Holland, 1979.

[Lakatos, 1976] Imre Lakatos. Proofs and refutations: the logic of mathematical discovery, Worrall, J. and Zahar, E.G. (eds), Cambridge University Press, 1976.

[Lamma *et al.*, 2000] E. Lamma, F. Riguzzi, L. M. Pereira. Belief revision by Lamarckian evolution. In *Applications of Evolutionary Computing: Procs. 1st European Workshop on Evolutionary Learning (EvoLEARN2001)*, E. J. W. Boers em et al., eds. pp. 404-416. Springer-Verlag, LNCS 2037, 2001.

[Lamma *et al.*, 2000a] Evelina Lamma, Luís Moniz Pereira and Fabrizio Riguzzi. Logic aided Lamarckian evolution. In P. Brazdil and R. Michalski (eds.), *Procs. Multi-Strategy Learning Workshop (MSL'00)*, Guimarães, Portugal, published by LIAAC - Porto University, June 2000.

[Lamma *et al.*, 2000a] Evelina Lamma, Fabrizio Riguzzi and Luís Moniz Pereira. Strategies in combined learning via logic programs. *Machine Learning*, 38(1/2): 63-87, January 2000.

[Leite and Pereira, 1997] João A. Leite and Luís Moniz Pereira. Generalizing updates: from models to programs. In ILPS'97 Workshop on Logic Programming and Knowledge Representation, LPKR'97, Port Jefferson 1997.

[Leite *et al.*, 2000] João A. Leite, Francisco C. Pereira, Amílcar Cardoso and Luís Moniz Pereira. Metaphorical mapping consistency via dynamic logic programming. In G. Wiggins (ed.), *Procs. AISB'00 Symposium on Creative and Cultural Aspects and Applications of AI and Cognitive Science*, pages 41-50, AISB, Birmingham, England, April 2000.

[Leite *et al.*, 2001] João A. Leite, José Júlio Alferes, Luís Moniz Pereira. Minerva - a dynamic logic programming agent architecture. In *Proc. Agent Theories, Architectures, and Languages (ATAL'01)*, John-Jules Meyer (ed.), Springer LNAI, 2001.

[Leite *et al.*, 2001a] João A. Leite, José Júlio Alferes and Luís Moniz Pereira. Multi-dimensional dynamic logic programming. In F. Sadri and K. Satoh (eds.), *Procs. CL-2000 Workshop on Computational Logic in Multi-Agent Systems (CLIMA'00)*, London, England, July 2000. An extended updated version appears as 'Multi-dimensional dynamic knowledge representation', In *Procs. 6th Int. Conf. on Logic Programming and Nonmonotonic Reasoning (LPNMR'01)*, T. Eiter and M. Truszczynski (eds), Springer LNAI 2001.

[Monteiro and Pereira, 1970] Luís Monteiro and Luís Moniz Pereira. Aspectos Cibernéticos da Epistemologia. In *Novas Perspectivas das Ciênciasdo Homem*, pp. 13-27, Biblioteca de Ciências Humanas 4, Editorial Presença, Lisboa, 1970.

[Nagel, 1997] Thomas Nagel. *The Last Word*, Oxford University Press, 1997.

[Newton, 1726] Isaac Newton. *Philosophiæ Naturalis Principia Mathematica*. Editio tertia & aucta emendata. Apud Guil & Joh. Innys, Regiæ Societatis typographos, 1726. Original quotation: *Corpus omne perseverare in statu suo quiescendi vel movendi uniformiter in directum, nisi quatenus illud a viribus impressis cogitur statum suum mutare.*

[Pereira, 1988] Luís Moniz Pereira. Inteligência Artificial: mito e ciência. In *Revista Colóquio-Ciências*, 3:1-13, October 1988, Fundação Calouste Gulbenkian, Lisboa.

[Pereira, 1999] Luís Moniz Pereira. The logical impingement of artificial intelligence, *Grazer Philosophische Studien (Internationale Zeitschrift Für Analytische Philosophie)*, Amsterdam/Atlanta, vol. 56, pp. 183-204, 1999.

[Pinker, 1997] Steven Pinker. *How the Mind Works*, W.W.Norton, 1997.

[Quaresma and Pereira, 1998] Paulo Quaresma and Luís Moniz Pereira. Modelling agent interaction in logic programming. In O. Barenstein (ed.), *Procs. of the 11th Int. Conf. on Applications of Prolog (INAP'98)*, Tokyo, Japan, September 1998.

[Quaresma and Rodrigues, 1999] Paulo Quaresma and Irene P. Rodrigues. A collaborative legal information retrieval system using dynamic logic programming. In *Procs. 7th Int. Conf. on AI and Law (ICAIL-99)*, pages 190-191. ACM SIGART, ACM Press 1999.

DOV M. GABBAY AND JOHN WOODS

FORMAL APPROACHES TO PRACTICAL REASONING: A SURVEY

1 INTRODUCTION

The purpose of this chapter is to indicate what kind of formal logical systems are suitable for the modelling of actual practical reasoning behaviour. The approach is to start from the familiar traditional notion of logic and indicate how more and more features need to be added to it. All such features arose during recent years through the attempts of the present authors as well as many of their colleagues to build better and better logical models of practical reasoning.

The urgent need for such models comes mainly from computer science and artificial intelligence. As there is pressure to build devices which help and/or replace the human in some of his daily activities, logic is called upon to provide realistic models of such activities and logic was therefore forced to evolve to accommodate human practice. There is also a benefit to some branches of traditional formal philosophy which have been studying aspects of human practical reasoning behaviour in the past 2000 years.

The structure of the chapter is such that it leads the reader from the traditional notion of a logic as a *consequence relation* to the more complex notion of what we call a *practical reasoning system*. These are the kinds of logical systems needed for a better modelling of practical reasoning. They actually evolved, as we have already said, from the authors' attempts to model various aspects of practical reasoning, as well as from attempts to formalise and unify the landscape of existing logical models put forward by the community.

In general, to specify a logical system in its broader sense we need to specify its components and describe how they relate to each other. Different kinds of logical systems have different kinds of components which bear different kinds of relationships to each other.

The following components are identified.

1. *The Language*
   This component simply defines our stock of predicates, connectives, quantifiers, labels, etc; all the kinds of symbols and syntactical structures involved in defining the basic components of the logic.

2. *Declarative Unit*
   This is the basic unit of the logic. In traditional logical systems (such as classical logic, modal logics, linear logic, etc.) the declarative unit

is simply a well formed formula. In more complex logics (such as *Labelled Deductive Systems*) it is a labelled formula or a database and a formula, etc.

3. *Databases*
This notion is that of a family of declarative units forming a *theory* representing intuitively the totality of our *assumptions*, with which we reason. In classical and modal logics this is a set of formulas. In linear logic it is a multiset of formulas. In the Lambek calculus it is a sequence of formulas. In *Labelled Deductive Systems* it is a structured labelled family of formulas.

   In general we need a notion that will include as a database a single declarative unit (compatibility criterion) and allow for more complex structures.

   Other notions need also be defined for databases. Among these are:

   - *Input* and *deletion*, i.e. how to add into and take out declarative units from a database. In classical and modal logic this is union and subtraction.
   - How to *substitute* one database Δ for a declarative unit φ inside another database containing φ. We need this notion to define *cut*.

     These notions are purely combinatorial and not logical in nature.[1]
   - Consistency and or acceptability of theories; integrity constraints.

4. *Consequence*
Now that we have the notion of a database, we can add a notion of consequence. In its simplest form it is a relation between two databases of the form $\Delta_1 \mid\!\sim \Delta_2$. $\mid\!\sim$ needs to satisfy some conditions. It means that Δ_2 *follows* in the logic from Δ_1. Of special interest is the notion $\Delta \mid\!\sim \varphi$, between a database Δ and a declarative unit φ.

   $\mid\!\sim$ can be specified set theoretically or semantically. Depending on its properties, it can be classified as monotonic or non-monotonic.

   As part of the notion of consequence we also include the notions of consistency and inconsistency.

[1] To explain what we mean, take the Lambek calculus. A declarative unit is any formula φ. A database is any sequence of formulas $\Delta = (\varphi_1, \ldots, \varphi_n)$. Input of φ into Δ can be either at the end of the sequence or the beginning, forming either $(\varphi_1, \ldots, \varphi_n, \varphi)$ or $(\varphi, \varphi_1, \ldots, \varphi_n)$. Similarly the corresponding deletion. Substitution for the purpose of Cut can be defined as follows:
The result of substituting $(\beta_1, \ldots, \beta_m)$ for α_i in $(\alpha_1, \ldots, \alpha_n)$ is the database $(\alpha_1, \ldots, \alpha_{i-1}, \beta_1, \ldots, \beta_m, \alpha_{i+1}, \ldots, \alpha_n)$.
Note that no logical notions are involved, only sequence handling.

5. *Proof theory, algorithmic presentation*
   One may give, for a given $|\!\sim$, an algorithmic system for finding whether $\Delta|\!\sim\varphi$ holds for a given Δ and φ. Such an algorithm is denoted by a computable metapredicate $S^{|\!\sim}(\Delta,\varphi)$. Different algorithmic systems can be denoted by further indices, i.e.

   $$S_1^{|\!\sim}(\Delta,\varphi),\ldots,S_i^{|\!\sim}(\Delta,\varphi).$$

   One view is to regard $S_i^{|\!\sim}$ as a mere convenience in generating or defining $|\!\sim$ and that the real logic is $|\!\sim$ itself.

   However, there are several established proof (algorithmic) methodologies that run across logics, and there are good reasons to support the view that at a certain level of abstraction we should consider any pair $(|\!\sim, S^{|\!\sim})$ as a logic. Thus classical logic via tableaux proofs is to be considered as a different logic from classical logic via Gentzen proofs.

6. *Mechanisms*
   The previous items (1)–(5) do not exhaust our list of components for a practical logical system. We need different mechanisms such as *abduction, revision, aggregation, actions*, etc. Such mechanisms make use of the specific algorithm S (of the logic $(|\!\sim, S^{|\!\sim})$) and define metalevel operations on a database. The particular version of such operations are considered as part of the logic. Thus a logic can be presented as $(|\!\sim, S^{|\!\sim}, S_{abduce}, S_{revise},\ldots)$.

   The notion of a database needs to be modified to include markers which can activate these mechanisms and generate more data. The language of the logic may include connectives that activate or refer to these mechanisms. Negation by failure is such an example.

7. Basic logical concepts such as metalevel object level distinctions, deduction theorems, relevance and resource limitations.

We shall see in later that it is convenient to present the metapredicate $S^{|\!\sim}$ as a three-place predicate. $S(\Delta,\varphi,x)$, where $x\in\{0,1\}$ or present it in the form $S(\Delta,\varphi)=x$. $S(\Delta,\varphi,1)$ means that the computation succeeds and $S(\Delta,\varphi,0)$ means that the computation finitely fails. The definitions of some of the $S_{mechanism}$ will make use of this fine tuning of the S predicate.

The rest of this chapter will motivate and exemplify the above notions.

2 LOGICAL SYSTEMS

We begin by presenting general answers to:

> *What is a logical system*?
> *What is a monotonic system*?
> *What is a non-monotonic system*?
> *What is a (formal) practical reasoning system*?

and related questions.

Imagine an expert system running on a personal computer, say the *Sinclair QL*. You put the data Δ into the system and ask it queries Q. We represent the situation schematically as:

$$\Delta?Q = \text{yes/no depending on the answer.}$$

We can understand the expert system because we know what it is supposed to be doing and we can judge whether its answers make reasonable sense. Suppose now that we spill coffee onto the keyboard. Most personal computers will stop working, but in the case of the QL, it may continue to work. Assume however that it now responds only to the symbol input and output; having lost its natural language interface. We want to know whether what we got is still 'logical' or not. We would not expect that the original expert system still works. Perhaps what we have now is a new system which is still a logic.

We are faced with the question of:

> *What is a logic*?

All we have is a sequence of responses:

$$\Delta_i?Q_i = \text{yes/no}$$

How do we recognize whether we have a logic at all?

This question was investigated by Tarski and Scott and they gave us an answer for monotonic logic systems. If we denote the relation

$$\Delta?Q = \text{yes} \qquad \text{by } \Delta \vdash Q$$

then this relation must satisfy three conditions to be a *monotonic logic*:

1. *Reflexivity*:
   $\Delta \vdash Q$ if $Q \in \Delta$.
2. *Monotonicity* :
   If $\Delta \vdash Q$ then $\Delta, X \vdash Q$.

3. *Transitivity (Lemma Generation, Cut)*:[2]
   If $\Delta \vdash X$ and $\Delta, X \vdash Q$ then $\Delta \vdash Q$.

To present a monotonic logic you have to mathematically define a relation '$\vdash$' satisfying conditions 1, 2, and 3. Such a relation is called a *Tarski consequence relation*. Non-monotonic consequence relations are obtained by restricting condition 2 to be the following condition:

2*. *Restricted Monotonicity:*
   If $\Delta \vdash Q$ and $\Delta \vdash X$ then $\Delta, X \vdash Q$.

This condition is discussed later in this section. In such a case we use the symbol '$\mid\!\sim$' instead of '$\vdash$'.

EXAMPLE 1. Let our language be based on atomic formulas and the single connective '$\Rightarrow$'. Define $\Delta \vdash_C Q$ to hold iff by doing classical truth tables for the formulas in Δ and Q, we find that whenever all elements of Δ get *truth*, Q also gets *truth*. We can check that conditions 1, 2, 3 hold. If so, we have therefore defined a logic. In this particular case, we also have an algorithm to check for a given Δ and Q, whether $\Delta \vdash Q$. In general, consequence relations can be defined mathematically without an algorithm for checking whether they hold or not.

EXAMPLE 2. For the same language (with '$\Rightarrow$' only) define $\vdash_I$ as the smallest set theoretical relation of the form $\Delta \vdash Q$ which satisfies conditions 1, 2, 3 together with the condition DT (*Deduction Theorem*):

DT: $\Delta \vdash A \Rightarrow B$ iff $\Delta \cup \{A\} \vdash B$.

We have to prove that this is a good definition. First notice that '$\vdash$' is a relation on the set $Powerset(Formulas) \times Formulas$ where $Formulas$ is the set of all formulas. Second we have to show that the smallest consequence relation '$\vdash$' required in the example does exist.

EXERCISE 3.

(a) Prove that $\vdash_I$ of the previous example 2 exists. It actually defines intuitionistic implication.

(b) Let $\Delta \vdash Q$ hold iff $Q \in \Delta$. Show that this is a monotonic consequence relation. (We call this civil servant logic, *Beamten Logik*)

(c) Similarly for $\Delta \vdash Q$ iff $\Delta = \{Q\}$. (This is the literally minded civil servant logic.)

---

[2]Cut has many versions. In classical logic they are all equivalent. In other logics they may not be. Here is another version:

4. *Second Version of Cut* :
   If $\Delta_1 \vdash X$ and $\Delta_2, X \vdash Q$ then $\Delta_1, \Delta_2 \vdash Q$.

We must be careful not to take a version of cut which collapses condition 2* to condition 2.

The difference between Examples 1 and 2 is that Example 2 does not provide us with an algorithm of when $\Delta \vdash_I Q$. The $\vdash_I$ is defined implicitly. For example how do we check whether (see Example 11.)

$$(((b \Rightarrow a) \Rightarrow b) \Rightarrow b) \Rightarrow a \vdash_I ?a$$

This motivates the need for algorithmic proof procedures.

We thus have the relationships depicted in Figures 1, 2 and 3.

Logics

Monotonic **Non-Monotonic**

Figure 1.

Large Monotonic Logics

Consequence Relation $\vdash_1$ defined mathematically in any manner.	**Consequence Relation** $\vdash_2$ defined mathematically in any manner.

Figure 2.

Consequence Relation $\vdash$

Algorithmic System $S_1^{\vdash}(\Delta, Q)$	**Algorithmic System** $S_2^{\vdash}(\Delta, Q)$

Figure 3.

$S_i^{\vdash}(\Delta, Q)$ is an algorithm for answering whether $\Delta \vdash Q$. Two properties are required:

Soundness: If $S_i^{\vdash}(\Delta, Q)$ succeeds then $\Delta \vdash Q$.

Completeness: If $\Delta \vdash Q$ then $S_i^{\vdash}(\Delta, Q)$ succeeds.

We assume that the algorithmic system is a recursive procedure for generating all pairs (with repetition) (Δ, Q) such that $\Delta \vdash Q$ holds.

There can be many algorithmic systems for the same logic. For example for classical logic, there are resolution systems, connection graph systems, Gentzen systems, semantic tableaux systems, Wang's method, etc.

An algorithmic system $S_i(\Delta, Q)$ may not be optimised in practice. It may be, for example, double exponential in complexity, etc. There are several heuristic ways of optimising it. If we try these optimising methods we get into different automated deduction systems for the algorithmic system S_i, denoted by: $O_1 S_i, O_2 S_i, \ldots$

In this case we require only soundness:

if $O_1 S_i(\Delta, Q)$ succeeds then $S_i(\Delta, Q)$ succeeds.

But we do not necessarily require completeness (i.e. the automated system may loop, even though the algorithmic system does not). (In fact, if the relation $\vdash$ is not recursively enumerable (RE), we may still seek an automated system. This will be expected to be only sound).

So far we were talking about monotonic systems. What is a non-monotonic system? Here we have a problem: Can we give conditions on $\Delta \mid\!\sim Q$, to characterize $\mid\!\sim$ as a non-monotonic system? (We use $\vdash$ for monotonic systems and $\mid\!\sim$ for non-monotonic systems). To seek this answer we must look at what is common to many existing non-monotonic systems. Do they have any common features, however weak these common features are? Before we proceed, let us give the main recognisable difference between monotonic and non-monotonic systems. Consider a database (1),(2), (3) and the query $?B$

(1)	$\neg A$	$?B$
(2)	$\neg A \Rightarrow B$	
(3)	*other data.*	

B follows from (1) and (2). It does not matter what the other data is. We do not need to survey the full database to verify that B follows. This is because of monotonicity. In non-monotonic reasoning, however, the deduction depends on the entire database. Thus if we put in more data, we get a new database and the deduction may not go through. Suppose we agree to list only positive atomic facts in the database. Then negative atomic facts are non-monotonically deduced simply from the fact that they are not listed. Thus a list of airline flights Vancouver–London which does list an 11:05-flight Monday to Saturday, would imply that there is no such flight on Sunday. Thus following this agreement, clause (1) of the database can be omitted provided the database is consistent, i.e. A is not listed in (3). We can deduce B by first deducing $\neg A$ (from the fact that it is not listed) and then deducing B from (2). To make sure A is not listed we must check the entire database.

Our original question was what are the conditions on $\mid\!\sim$ to make it into a non-monotonic logic?

We propose to replace condition 2. on $\vdash$ of monotonicity by condition 2* already mentioned, namely:

2*. *Restricted Monotonicity*:
If $\Delta \mid\!\sim X$ and $\Delta \mid\!\sim Q$ then $\Delta, X \mid\!\sim Q$.

Its meaning is that if X, Q are expected to be true by Δ (i.e. $\Delta \mid\!\sim X$ and $\Delta \mid\!\sim Q$) then if X is actually assumed true, then Q is still expected to be true (i.e. $\Delta, X \mid\!\sim Q$).

In Section 3 we will come back to the question of finding a right definition for non-monotonic consequence relation. We will see in that context that it might be interesting to replace also the cut by weaker rules. We shall further claim that it is best to regard the term 'non-monotonic' as expressing the absence of monotonicity and that many different features fall under this umbrella, among them resource considerations. Some of these features are orthogonal and complementary to each other.

Given a non-monotonic system $\mid\!\sim$, we can still ask for algorithmic systems $S_i^{\mid\sim}(\Delta, Q)$ for $\mid\!\sim$. In the non-monotonic case these are rare. Most non-monotonic systems are highly non-constructive and are defined using complex procedures on minimal models or priority of rules or non-provability considerations. So the metatheory for $\mid\!\sim$ is not well developed. There is much scope for research here. Later on we will introduce the notion of *Labelled Deductive Systems* (*LDS*) which will yield in a systematic way a proof theory for many non-monotonic systems.

We are now ready to answer provisionally the question of what is a logical system. We propose a first answer which may need to be modified later on.

DEFINITION 4 (Logical System version 1). A logical system is a pair $(\Vdash, S^{\Vdash})$, where $\Vdash$ is a consequence relation (monotonic or non-monotonic, according to whatever definition we agree on) and $S^{\Vdash}$ is an algorithmic system for $\Vdash$. $S^{\Vdash}$ is sound and complete for $\Vdash$.[3]

Thus different algorithmic systems for the same consequence relation give rise to different logical systems. So classical logic presented as a tableau system is not the same logic as classical logic presented as a Hilbert system.

Here are some examples of major proof systems:

- Gentzen
- Tableaux
- Semantics (effective truth tables)
- Goal directed methodology

[3] Of course we also need many additional notions such as the notion of what is a theory (database), input into a theory, consistency/inconsistency, etc.

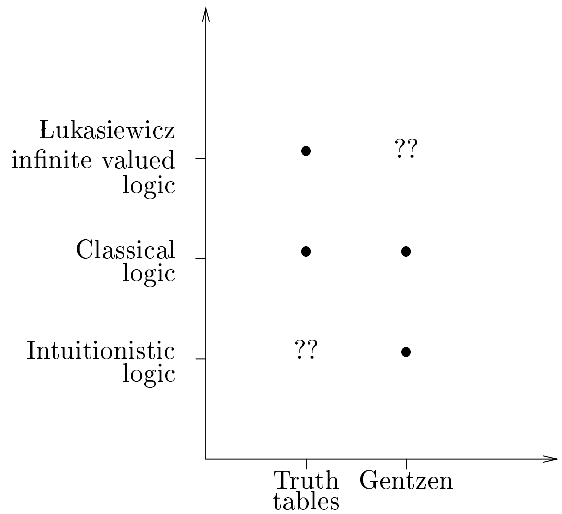

Figure 4. Logics landscape

- Resolution
- Labelled Deductive systems, etc.

Definition 4 is supported by the following two points:

1. We have an intuitive recognition of the different proof methodologies and show individual preferences to some of them depending on the taste and the need for applications.
2. Slight variations in the parameters of the proof systems can change the logics significantly.
3. Better fit to practical reasoning daily practice. There are many agreed procedures as part of our daily life and these are better modelled in an algorithmic environment.

Figure 4 is an example of such a relationship.

In the truth table methodology classical logic and Łukasiewicz logic are a slight variation of each other. They resemble intuitionistic logic less closely. In the Gentzen approach, classical and intuitionistic logics are very close while Łukasiewicz logic is a problem to characterise.

This evidence suggests strongly that the landscape of logics is better viewed as a two-dimensional grid.

The reader may ask why we need $\Vdash$, if we have $S^{\Vdash}$ from which $\Vdash$ can be obtained? The answer is that the definition of $\Vdash$ is needed. $\Vdash$ is introduced via a mathematical definition which reveals the intended meaning of $\Vdash$ as

separate from the algorithmic means of computing it. So by saying a logical system is a pair $(\Vdash, S^{\Vdash})$ we do not intend only a set theoretical definition of $\Vdash$ and an algorithm $S^{\Vdash}$, but also an expression of the intended meaning for $\Vdash$ as well.

The reader should note that Definition 4 is a central definition and a serious departure from current conceptual practice. It will be properly motivated throughout these chapters. Let us use the notation $\tau = (\Vdash_\tau, S^{\Vdash}_\tau)$ for a logical system τ. $\Vdash_\tau$ is its consequence relation and $S^{\Vdash}_\tau$ is an algorithmic system for $\Vdash_\tau$. We also note that in case $\Vdash_\tau$ is not RE, Recursively Enumerable, (as may happen often in non-monotonic logics) we will be satisfied with $S^{\Vdash}_\tau$ which is only sound for $\Vdash$.

Here is a further example:

EXAMPLE 5 (Modal logic **S4**).

1. Consider a language with atoms p, q, r, ...the classical connectives $\neg$, $\wedge$ and the unary connective $\Box$. Let h be a function assigning to each atom a set of points in the Euclidean plane $\mathcal{R}^2$. Let:

   - $h(A \wedge B) = h(A) \cap h(B)$.
   - $h(\neg A) =$ complement of $h(A)$.
   - $h(\Box A) =$ topological interior of $h(A)$.

   Let $\models A =_{df} \forall h[h(A) = \mathcal{R}^2]$.

   Let $A_1, \ldots, A_n \models B$ iff $\forall h(\bigcap_i h(A_i) \subseteq h(B))$.

   Then $\models$ is a consequence relation.

   Let $A \Rightarrow B$ be defined as $\neg(A \wedge \neg B)$ and let $A \leftrightarrow B$ be defined as $(A \Rightarrow B) \wedge (B \Rightarrow A)$. Then we have for example: $\models \Box(A \wedge B) \Leftrightarrow \Box A \wedge \Box B$.

2. Let * be a translation from the previous language into classical logic. For each atom q_i associate a unary predicate $Q_i(t)$, with one free variable t. Let R be a binary relation symbol. Translate as follows (note that the translation function depends of t):

   - $(q_i)^*_t = Q_i(t)$.
   - $(A \wedge B)^*_t = A^*(t) \wedge B^*(t)$.
   - $(\neg A)^*_t = \neg A^*(t)$.
   - $(\Box A)^*_t = \forall s(tRs \Rightarrow (A)^*_s)$.

   Let $A_1, \ldots, A_n \Vdash A$ hold iff in predicate logic one can prove:

   $$\textit{classical logic} \vdash [\forall x(xRx) \wedge \forall xyz(xRy \wedge yRz \Rightarrow xRz)] \Rightarrow \forall t(\bigwedge_i (A_i)^*_t \Rightarrow (A)^*_t).$$

The two consequence relations $\models$ and $\Vdash$ are defined in a completely different way. They are the same, however, from the mathematical point of view. i.e. $\Delta \Vdash A$ iff $\Delta \models A$ holds. Their meaning is not the same.

To define an algorithmic system for $\Vdash$ or $\models$ we can modify and use any theorem prover for classical logic.

3. It is possible to give a Hilbert formulation for this consequence relation with the following axioms together with modus ponens (necessitation is derivable):

   (a) $\Box A$, where A is an instance of a truth functional tautology.

   (b) $\Box(\Box(A \Rightarrow B) \Rightarrow (\Box A \Rightarrow \Box B))$

   (c) $\Box(\Box A \Rightarrow \Box\Box A)$

   (d) $\Box A \Rightarrow A$

   (e) $\Box(\Box A \Rightarrow A)$

We have $A_1, \ldots, A_n \vdash B$ iff $\bigwedge_i A_i \Rightarrow B$ is a theorem of the Hilbert system.

This example illustrates our point that even with the same S, $(\models, S)$ and $(\Vdash, S)$ as defined are *not the same logic*!

Let us revisit monotonicity's three levels of presentation. These were:

Mathematical definition of $\vdash$	$\Delta \vdash Q$
Algorithmic procedures (recursively enumerable)	$S_i^{\vdash}(\Delta, Q)$
Optimising automated systems (polynomial time)	$OS_i^{\vdash}(\Delta, Q)$

Experience shows that if we take $S_i^{\vdash}$ and change the computation slightly, to S_i^*, we may get another logic. For example $S(\Delta, Q)$ is an algorithmic system for intuitionistic logic. Change S a little bit to S^*, and $S^*(\Delta, Q)$ gives you classical logic. These type of connections are very widespread, to the extent that we get a much better understanding of a logic $\vdash$, not only through its own algorithmic systems, but also (and possibly even chiefly) through its being a 'changed' version of another automated system for another logic.

It is difficult to appreciate fully what is being said now because we are talking in the abstract, without examples — we will consider some examples later on. There is no choice but to talk abstractly *in the beginning*.

Another important point is the role of failure. Assume we have an algorithmic system for some logic. The algorithmic system can be very precise, in fact, let us assume it is an automated system. Suppose we want to ask $\Delta?Q$. We can look at the rules of the automated system and see immediately that it is looping (a loop checker is needed; we can record the history

of the computation and be able to detect that we are repeating ourselves) or possibly finitely failing (i.e. we try all our computation options and in each case we end up in a situation where no more moves are allowed). We add new connectives to the logic, denoted by $loop(Q)$ and $fail(Q)$ and write formulas like $loop(Q) \Rightarrow R$ or $fail(Q) \Rightarrow P$. This is similar to Prolog's negation by failure. We get *new* non-monotonic logics out of old monotonic logics. Here we are connecting the monotonic hierarchy of logics with the non-monotonic one. The abduction mechansim can also be viewed in this way, as $fail(q) \Rightarrow q$.

Our final point here concerns the general nature of theorem proving (or automated reasoning). So far we talked about $S_i^{\vdash}(\Delta, Q)$. All our automated rules have the form e.g. $S_i^{\vdash}(\Delta, Q \Rightarrow R)$ if $S_i^{\vdash}(\Delta \cup \{Q\}, R)$; in other words, the success of $\Delta \vdash Q \Rightarrow R$ reduces to that of $\Delta, Q \vdash R$ *in the same logic* $\vdash$. This suggests the need for taking the logic as a *parameter*, in view of the fact that we get things like $S(\Delta, Q, \text{logic } 1)$ if $S(\Delta', Q', \text{logic } 2)$.

We thus have automated system defining several logics by mutual recursion.

3 EXAMPLES OF LOGICAL SYSTEMS

This section discusses a number of examples of logical systems. We begin with the definition of a Kripke model which we shall use to define several logics. Note that a consequence relation can be presented via several completely unrelated semantical interpretations. A single semantical interpretation can be slightly changed to give rise to different logics.

DEFINITION 6.

1. A Kripke model (structure) for modal logic has the form $\mathbf{m} = (T, R, a, h)$, where T is a set of possible worlds, $a \in T$ is the actual world and $R \subseteq T \times T$ is the accessibility relation. h is an assignment giving for each atomic q a subset $h(q) \subseteq T$.

   h can be extended to all wffs as follows:

   - $h(\neg A) = T - h(A)$
   - $h(A \wedge B) = h(A) \cap h(B)$
   - $h(\Box A) = \{t |$ for all s, if tRs then $s \in h(A)\}$
   - We say $\mathbf{m} \models A$ iff $a \in h(A)$

   The following holds for the logic $\mathbf{K}$:

   - $\mathbf{K} \models A$ iff for all $\mathbf{m}, \mathbf{m} \models A$

2. The modal logics **T**, **B**, **K4**, **S4**, **S5** are defined by the class of all Kripke models where the accessibility relation R takes the following interesting properties:

   T Reflexivity

   B Symmetry and Reflexivity

   K4 Transitivity

   S4 Reflexivity and transitivity

   S5 Reflexivity, transitivity and symmetry

3. Kripke models can be used to characterise intuitionistic logic as well ($\vdash$ of Example 2) as follows:

   Consider models of the form (T, R, a, h) as above, with R reflexive and transitive and h satisfying the following persistence condition for all q

   - $t \in h(q)$ and tRs imply $s \in h(q)$.

   We now associate with each wff A two subsets $[A]^{Beth}$ and $[A]^{Kripke}$ as follows:

   - $[q]^{Kripke} = h(q)$
   - $[q]^{Beth} = \{x \in T|$ every maximal chain Π through x intersects $h(q)\}$, where a maximal chain π through x is a maximal R linearly ordered subset of T containing x. In this case we say that $h(q)$ *bars* x.
   - $[A \wedge B]^{Beth} = [A]^{Beth} \cap [B]^{Beth}$
   - $[A \wedge B]^{Kripke} = [A]^{Kripke} \cap [B]^{Kripke}$
   - $[A \Rightarrow B]^{Beth} = \{x \in T|$ for all y, tRy and $y \in [A]^{Beth}$ imply $y \in [B]^{Beth}\}$

     Similarly for $[A \Rightarrow B]^{Kripke}$

   - $[\neg A]^{Beth} = \{x|$ for all y, xRy implies $y \notin [A]^{Beth}\}$

     Similarly for $[\neg A]^{Kripke}$

   - $[A \vee B]^{Kripke} = [A]^{Kripke} \cup [B]^{Kripke}$
   - $[A \vee B]^{Beth} = \{x|$ every maximal chain through x intersects $[A]^{Beth} \cup [B]^{Beth}\}$
   - Define $\mathbf{m} \models_{Beth} A$ iff $[A]^{Beth} = \mathrm{T}$
     $\mathbf{m} \models_{Kripke} A$ iff $[A]^{Kripke} = \mathrm{T}$

- Define consequence relations $\Vdash_{Beth}$ and $\Vdash_{Kripke}$ by:

  $A_1, \ldots, A_n \Vdash_{Beth} B_1, \ldots, B_k$ iff for each model $\mathbf{m}$ there exists B_j such that $\mathbf{m} \Vdash_{Beth} \bigwedge A_i \Rightarrow B_j$

  Similarly for $\Vdash_{Kripke}$.

EXAMPLE 7. Consider a language with $\wedge$, $\Rightarrow$, $\neg$, $\vee$ and $\Box$. Define a Hilbert system for a modal logic $\mathbf{K}$ for modality $\Box$. $\mathbf{K}$ has the axioms and rules as follows:

1. Any instance of a truth functional tautology.
2. $\Box(A \Rightarrow B) \Rightarrow (\Box A \Rightarrow \Box B)$
3. The rules:

$$\frac{\vdash A, \vdash A \Rightarrow B}{\vdash B} \quad \text{and} \quad \frac{\vdash A}{\vdash \Box A}$$

Define $A_1, \ldots, A_n \vdash_{\mathbf{K}} A$ iff $\vdash \bigwedge A_i \Rightarrow A$. It is complete for all Kripke structures.

Consider the additional condition on Kripke structures that only the actual world is reflexive (i.e. we require aRa but not generally $\forall x(xRx)$). The class of all models with aRa defines a new logic). Call this logic $\mathbf{K1}$. We cannot axiomatise this logic by adding the axiom $\Box A \Rightarrow A$ to $\mathbf{K}$ because this will give us the logic $\mathbf{T}$ (complete for reflexivity of every possible world, not just the actual world). We observe, however, that $\vdash_{\mathbf{K}} A$ implies $\vdash_{\mathbf{K1}} \Box A$. If we adopt the above rule together with $\vdash_{\mathbf{K1}} \Box A \Rightarrow A$, we will indeed get, together with modus ponens an axiomatization of $\mathbf{K1}$.

The axiomatization of $\mathbf{K1}$ is obtained as follows:

Axioms:

1. Any substitution instance of a theorem of $\mathbf{K}$.
2. $\Box A \Rightarrow A$.

Rules: Modus ponens only.

We can define an extension of $\mathbf{K1}$, (call it $\mathbf{K1}_{[2]}$), by adding a $\Box^2$ necessitation rule

$$\frac{\vdash A}{\vdash \Box^2 A}.$$

EXAMPLE 8 (Classical logic with restart). The following is an algorithmic (possibly a phantom algorithms) presentation of classical propositional logic.

(See [Gabbay, 1998] for more details). We choose a formulation of classical logic using $\wedge, \to, \bot$. We write the formulas in a 'ready for computation' form as clauses as follows:

1. q is a clause, for q atomic, or $q = \bot$
2. If $B_j = \bigwedge_i A_{ij} \to q_i$ are clauses for $j = 1, \ldots, k$ then so is $\bigwedge_j B_j \to q$, where q is atomic or $q = \bot$.
3. It can be shown that every formula of classical propositional logic is classically equivalent to a conjunction of clauses.

We now define a goal directed computation for clauses.

4. Let $S(\Delta, G, H)$ mean the clause G succeeds in the computation from the set of clauses Δ given the history H, where the history is a set of atoms or $\bot$.
   (a) $S(\Delta, q, H)$ if $q \in \Delta$ for q atomic or $\bot$.
   (b) $S(\Delta, q, H)$ if $\Delta \cap H \neq \varnothing$.
   (c) $S(\Delta, \bigwedge_j(\bigwedge_i A_{ij} \to q_j) \to q, H)$ if $S(\Delta \cup \{\bigwedge_i A_{ij} \to q_j\}, q, H)$.
   (d) $S(\Delta \cup \{\bigwedge_{j=1}^{k}(\bigwedge_{i=1}^{m_j} A_{ij} \to q_j) \to q)\}, q)$ if $\bigwedge_j S(\Delta \cup \{A_{ij} \mid i = 1, \ldots, m_j\}, q_j, H \cup \{q\})$
5. The computation starts with $S(\Delta, G, \varnothing)$, to check whether $\Delta \vdash G$.

We have the following theorem:

$$S(\Delta, q, H) \text{ iff } \Delta \vdash q \vee \bigvee H.$$

EXERCISE 9. Consider the language with $\Rightarrow$ and $\neg$ and the many valued truth tables below. Define:

$$A_1, \ldots, A_n \vdash B \text{ iff } \quad A_1 \Rightarrow (A_2 \Rightarrow \ldots (A_n \Rightarrow B) \ldots)$$
gets always value 1 under all assignments

What consequence relation do we get? What logic is it?

Table 1.

$\Rightarrow$	1	1/2	0
1	1	1/2	0
1/2	1	1	1/2
0	1	1	1

	$\neg$
1	0
1/2	1/2
0	1

EXERCISE 10. Define a Horn clause with negation by failure as any wff of the form $\bigwedge a_i \wedge \bigwedge \neg b_j \Rightarrow q$, where a_i, b_j, q are atomic; q is called the head of the clause. Define a computation as follows:

Table 2.

$\Rightarrow$	1	2	3	4
1	1	2	3	4
2	1	1	3	3
3	1	2	1	2
4	1	1	1	1

	$\neg$
1	4
2	3
3	2
4	1

1. $\Delta ?q = success$ if $q \in \Delta$.
2. $\Delta ?q = failure$ if q is not head of any clause in Δ.
3. $\Delta ?q = success$ if for some $\bigwedge a_i \wedge \bigwedge \neg b_j \Rightarrow q$ in Δ we have that $\Delta ?a_i = success$ for all i and $\Delta ?\neg b_j = success$ for all j.
4. $\Delta ?\neg b = success$ (respectively failure) if $\Delta ?b = failure$ (respectively success).
5. $\Delta ?q = failure$ if for each clause $\bigwedge a_i \wedge \bigwedge \neg b_j \Rightarrow q \in \Delta$ we have either for *some* i, $\Delta ?a_i = failure$ or for some j, $\Delta ?\neg b_j = failure$.

(a) Show that propositional Horn clause Prolog with negation by failure is a non-monotonic system, i.e. if we define $\Delta \mid\!\sim Q$ iff (definition) Q succeeds in Prolog from data Δ, then $\mid\!\sim$ is a non-monotonic consequence relation according to our definition. (To show (3) and (2*) assume X is positive.)

(b) *(Challenge)* Prove in Prolog $\mid\!\sim$ that:

$$\text{if } \Delta \mid\!\sim q \text{ and } \Delta, d \mid\!\sim \neg q \text{ then } \Delta \mid\!\sim \neg d \text{ (for } d \text{ atomic).}$$

Similarly,

$$\text{if } \Delta \mid\!\sim \neg q \text{ and } \Delta, d \mid\!\sim q \text{ then } \Delta \mid\!\sim \neg d.$$

EXAMPLE 11. Recall the definition of $\Phi \vdash_I A$ as the smallest Tarski relation on the language with $\Rightarrow$ such that the equation DT holds.

DT: $\Phi \vdash A \Rightarrow B$ iff $\Phi, A \vdash B$.

According to Exercise 3, $\vdash_I$ exists.

Our algorithmic problem is how do we show for a given $\Phi \vdash_I ?A$ whether it holds or not.

Take the following (Hudelmaier):

$$(((b \Rightarrow a) \Rightarrow b) \Rightarrow b) \Rightarrow a \vdash_I ?a$$

We can only use the means at our disposal, in this case DT. Certainly, for cases of the form $\Phi \vdash_I A \Rightarrow B$ we can reduce to $\Phi, A \vdash_I B$. This is a sound policy, because we simplify the query and monotonically strengthen the data at the same time.

When the query is atomic we cannot go on. So we must look for patterns.

Out of a, b we can make the following possible formulas with at most one $\Rightarrow$:

$a \Rightarrow b$
a
b
$b \Rightarrow a$
$a \Rightarrow a$
$b \Rightarrow b$

The following are possible consequences:

$a, b \vdash ?a \Rightarrow b$
$a, a \Rightarrow b \vdash ?b$
$a \Rightarrow b \vdash ?a \Rightarrow b$
$a \Rightarrow a \vdash ?b \Rightarrow b$

Shuffling around and recognizing cases of reflexivity we can get:

1. $a, a \Rightarrow b \vdash b$ (from reflexivity and DT)

2. $\vdash a \Rightarrow a$

3. $b \vdash a \Rightarrow b$

Back' to our example:

$(((b \Rightarrow a) \Rightarrow b) \Rightarrow b) \Rightarrow a$	$\vdash_I ?a$
same	$\vdash_I ?((b \Rightarrow a) \Rightarrow b) \Rightarrow b$
add $(b \Rightarrow a) \Rightarrow b$	$\vdash_I ?b$
same	$\vdash_I ?b \Rightarrow a$
add b	$\vdash_I ?a$
same	$\vdash_I ?((b \Rightarrow a) \Rightarrow b) \Rightarrow b$
same	$\vdash_I ?b$ (*success*: you already have b)

If we look at the kind of applications studied in this handbook we see that we need a more refined notion of a logical system, to enable us to cope with the needs of the applications. The following sections survey our options.

4 STRUCTURED CONSEQUENCE

The next move in the notion of a logical system is to observe that part of the logic must also be the notion of what the logic accepts as a theory. Theories have structure, they are not just sets of wffs. They are structures of wffs. Different notions of structures give rise to different logics, even though the logics may share the same notion of consequence for individual wffs.

We need the following:

1. A notion of structure to tell us what is to be considered a theory for our logic. For example, a theory may be a multiset of wffs or a list of wffs or a more general structure. The best way to define the general notion of a structure is to consider models M of some classical structure theory τ and a function $\mathbf{f}: M \mapsto$ wffs of the logic. A theory Δ is a pair $(M, \mathbf{f})$. We write $t : A$ to mean $\mathbf{f}(t) = A$.

2. A notion of insertion into the structure and deletion from the structure. I.e. if $t \in M$ and $s \notin M$ we need to define $M' = M + \{s\}$ and $M'' = M - \{t\}$. M' and M'' must be models of τ. When theories were sets of wffs, insertion and deletion presented no problems. For general structures we need to specify how it is done.

3. A notion of substitution of one structure M_1 into another M_2 at a point $t \in M_2$. This is needed for the notion of cut. If $(M_2, \mathbf{f}_2) \mid\!\sim A$ and for some $t \in M_2, (M_1, \mathbf{f}_1) \mid\!\sim \mathbf{f}_2(t)$, we want to 'substitute' '$(M_1, \mathbf{f}_1)$' for 't' in M_2 to get $(M_3, \mathbf{f}_3)$ such that $(M_3, \mathbf{f}_3) \mid\!\sim A$.

EXAMPLE 12. Let the structure be lists. So let, for example, $\Delta = (A_1, \ldots, A_n)$. We can define $\Delta + A = (A_1, \ldots, A_n, A)$ and $\Delta - \{A_i\} = (A_1, \ldots A_{i-1}, A_{i+1}, \ldots, A_n)$. Substitution of $\Gamma = (B_1, \ldots, B_k)$ for place i in Δ (replacing A_i) gives $\Delta[i/\Gamma] = (A_1, \ldots, A_{i-1}, B_1, \ldots, B_k, A_{i+1}, \ldots, A_n)$.

The cut rule would mean

- $\Delta \mid\!\sim C$ and $\Gamma \mid\!\sim A_i$ imply $\Delta[i/\Gamma] \mid\!\sim C$.

We need now to stipulate the minimal properties of a structured consequence relation. These are the following:

Identity $\quad \{t : A\} \mid\!\sim t : A$

Surgical Cut $\dfrac{\Delta \mid\!\sim t : A; \Gamma[t : A] \mid\!\sim s : B}{\Gamma[t/\Delta] \mid\!\sim r : B}$

Typical good examples of structured consequence relations are algebraic labelled deductive systems based on implication $\to$.

The basic rule is modus ponens.

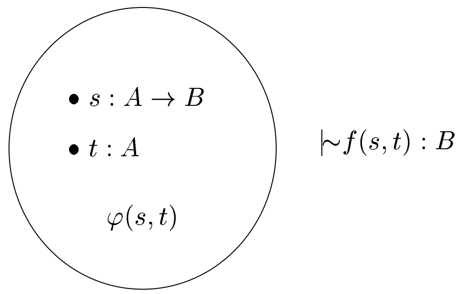

Figure 5.

- $$\frac{s : A \to B; t : A; \varphi(s,t)}{f(s,t) : B}$$

t, s are labels, φ is a compatibility relation on labels and $f(s,t)$ is the new label.

In terms of databases the rule means as in Figure 5

The labels can be resource, time, relevance, strength/reliability, complete file. See [Gabbay, 1996].

5 ALGORITHMIC STRUCTURED CONSEQUENCE RELATION

Just as in the case of ordinary consequence relations, different algorithms on the structure are considered as different logics. The general presentation form of most (maybe all) algorithms is through a family of rules of the form

- $\Delta \mathrel{|\!\sim} ?\Gamma, \Pi$ reduces to $\Delta_i \mathrel{|\!\sim} ?\Gamma_i, \Pi_i, i = 1, \ldots, n$;

if we use a traditional method of display we will write:

- $$\frac{\Delta_1 \mathrel{|\!\sim} \Gamma_1, \Pi_1; \ldots; \Delta_n \mathrel{|\!\sim} \Gamma_n, \Pi_n}{\Delta \mathrel{|\!\sim} \Gamma, \Pi}.$$

We need the notion of a formula $t : A$ in the structure Π *being used* in the rule. We further need to have some complexity measure available which decreases with the use of each rule. $\Delta, \Gamma, \Delta_i, \Gamma_i$ are structured theories and Π, Π_i are parameters involved in the algorithm, usually the history of the computation up to the point of application of the rule. There may be side conditions associated with each rule restricting its applicability.

Some rules have the form

- $\Delta \mathrel{|\!\sim} ?\Gamma, \Pi$ reduces to *success*

or in a traditional display

- $\overline{\Delta \mid\!\sim \Gamma, \Pi}$.

These are axioms.

Having the rules in this manner requires additional (decidable) metapredicates on databases and parameters. We need the following:

1. $\Psi_0(\Delta, \Gamma, \Pi)$ recognising that $\Delta \mid\!\sim \Gamma, \Pi$ is an axiom.

2. $\Psi_1(\Delta, \Gamma, \Pi)$ recognising that $\Delta \mid\!\sim \Gamma, \Pi$ is a failure (no reduction rules apply).

3. $\Psi_2(\Delta, \Delta_1, \Delta_2, \Pi, \Pi_1, \Pi_2)$ says that Δ is the result of inserting Δ_2 into Δ_1 (also written as $\Delta = \Delta_1 + \Delta_2$, ignoring the parameters).

4. $\Psi_4^n(\Delta, \Delta_1, \ldots, \Delta_n, \Pi, \Pi_1, \ldots, \Pi_n)$ says that Δ is decomposed into $\Delta_1, \ldots, \Delta_n$. The decomposition is not necessarily disjoint. So for example, the reduction rule $\Delta \mid\!\sim ?\Gamma$ if $\Delta_i \mid\!\sim ?\Gamma_i, i = 1, \ldots, n$ may require that $\Psi_4^n(\Delta, \Delta_1, \ldots, \Delta_n)$ and $\Psi_4^n(\Gamma, \Gamma_1, \ldots, \Gamma_n)$ both hold, (ignoring the parameters).

There may be more Ψs involved.

The Ψs may be related. For example $\Psi_4^n(\Delta, \Delta_1, \ldots, \Delta_n)$ may be $\Delta = (\Delta_1 + (\Delta_2 + (\ldots + \Delta_n) \ldots)$.

The reader should note that with structured databases, other traditional notions associated with a logic change their relative importance, role and emphasis. Let us consider the major ones.

Inconsistency

In traditional logics, we reject inconsistent theories. We do not like having both A and $\neg A$ among the data. In structured databases, we have no problem with that. We can have $\Delta = \{t : A, s : \neg A\}$ and what we prove from Δ depends on the logic. Even when we can prove everything from a database, we can use labels to control the proofs and make distinctions on how the inconsistency arises. A new approach to inconsistency is needed for structured databases. Inconsistency has to be redefined as *acceptability.* Some databases are not acceptable. They may be consistent or inconsistent. Consistency is not relevant, what is relevant is their acceptability.

We have, for example, the notion of integrity constraints in logic and commercial databases. We may have as an integrity constraint for a practical database that it must list with each customer's name also his telephone number. A database that lists a name without a telephone number may be consistent but is unacceptable. It does not satisfy integrity constraints. See [Gabbay and Hunter, 1991] and [Gabbay and Woods, 2001a] for more discussion.

Deduction theorem

If a theory Δ is a set of sentences we may have for some A, B that $\Delta \not\mathrel{|\!\sim} B$ but $\Delta \cup \{A\} \mathrel{|\!\sim} B$. In which case we can add a connective $\twoheadrightarrow$ satisfying $\Delta \mathrel{|\!\sim} A \twoheadrightarrow B$ iff $\Delta \cup \{A\} \mathrel{|\!\sim} B$.

If Δ is a structured database, we have an insertion function $\Delta + A$, adding A into Δ and a deletion function, $\Delta - A$, taking A out of Δ. We can stipulate that $(\Delta + A) - A = \Delta$.

The notion of deduction theorem is relative to the functions $+$ and $-$. Let $\twoheadrightarrow_{(+,-)}$ be the corresponding implication, then we can write $\Delta + A \mathrel{|\!\sim} B$ iff $(\Delta + A) - A \mathrel{|\!\sim} A \twoheadrightarrow_{(+,-)} B$.

We can have many $\twoheadrightarrow$s and many deduction theorems for many options of $(+, -)$ pairs.

Take, for example, the Lambek calculus.

The databases are lists, $(A_1, \ldots, A_n)$ of wffs. We can have two sets of $+$ and $-$, namely we can have $+_r, -_r$ (add and delete from the right hand side) and $+_l, -_l$ (add and delete from the left). This gives us two arrows and two deduction theorems

$$(A_1, \ldots, A_n) \mathrel{|\!\sim} A \to_r B \text{ iff } (A_1, \ldots, A_n, A) \mathrel{|\!\sim} B$$

and

$$(A_1, \ldots, A_n) \mathrel{|\!\sim} A \to_l B \text{ iff } (A, A_1, \ldots, A_n) \mathrel{|\!\sim} B.$$

In fact the smallest bi-implicational logic with $\to_r, \to_l$ only and databases which are lists which satisfy the two deduction theorems is the Lambek calculus.

Skolemisation and substitution

Consider $t : \exists x A(x)$. The label can be time, context, resource, etc. We want to Skolemise and substitute a constant c for the existential quantifier. It makes sense to annotate the constant with the label of where it was created. Thus we create c^t and write the rule

$$\frac{t : \exists x A(x)}{t : A(c^t)}.$$

c^t certainly exists at label t. We write this as $t : c^t$. Let s be another label (say s is accessible to t according to some relation $\rho(t, s)$). ρ may be a possible world accessibility relation, a resource comparison relation or whatever goes between labels. We may wish to state that constants available at label t are also available at s. We write the following *visa rule*:

$$\frac{t : c; \rho(t, s)}{s : c}$$

EXAMPLE 13. Assume the labels are worlds and ρ is accessibility. Let us see if we can show that $\exists x \Diamond A(x) \vdash \Diamond \exists x A(x)$

1. $t : \exists x \Diamond A(x)$ assumption
2. $t : \Diamond A(c^t)', t : c^t$ Skolemization
3. $\rho(t, s); s : A(c^t)$ rule for $\Diamond$, allowing us to find a possible world where $A(c^t)$ holds.
4. We cannot deduce $s : \exists x A(x)$ because we do not know whether c^t exists at s. We need a *visa rule* to that effect. If this is available we can get $s : \exists x A(x)$
5. $t : \Diamond \exists x A(x)$, from 4.

Another possibility afforded by using labels is to restrict the range of substitution of quantifiers and free variables. Given $t : A(x)$, x a free variable traditional logic allows the substitution of any term in it. We saw that we can restrict the substitution $\theta_t(x)$ only to c such that $t : c$ holds (existing at t). But we can have rules governing the substitution as well, e.g.

$$\frac{t : c; t : \theta}{c \in \theta}$$

$$\frac{t : \theta; \rho(t, s), s : \theta'}{\theta \subseteq \theta'}$$

and so on.

Such rules are important in applications of linguistics.

6 MECHANISMS

Our notion of logical systems so far is of pairs $(\mid\!\sim, S^{\mid\sim})$, where $\mid\!\sim$ is a structured consequence relation between structured databases and $S^{\mid\sim}$ is an algorithm for computing $\mid\!\sim$. The data items in any database $\Delta = (M, \mathbf{f})$ are wffs, i.e. for $t \in M, \mathbf{f}(t)$ is a wff. We know from many applications that items residing in the database need not be the data itself but can be mechanisms indicating how to obtain the data. Such mechanisms can be sub-algorithms which can be triggered by the main $S^{\mid\sim}$ algorithm or in the most simple case just a link to another database.

We regard such mechanisms as part of the logic. They are ways of extending our databases Δ with more data without having to put them explicitly into the structure of Δ.

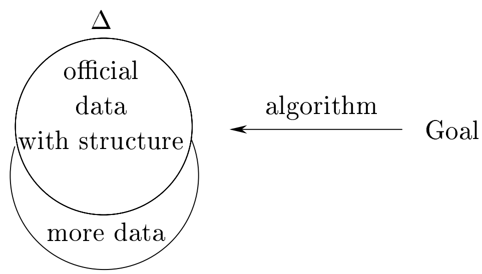

Figure 6. Mechanisms extend data

Among the well known mechanisms are

- abduction
- default
- other non-monotonic mechanisms (circumscription, negation as failure, etc).
- update and revision.

Figure 6 shows what a database looks like. Formally a database $\Delta = (M, \mathbf{f})$ can be such that for $t \in M, \mathbf{f}(t)$ is a mechanism. Of course these mechanisms depend on $S^{|\!\sim}$.

These mechanisms are well studied in the literature but the traditional way of perceiving them is that they depend on $|\!\sim$ and not necessarily on the proof method ($S^{|\!\sim}$). Our view is that

1. they are part of Δ;

and that

2. they depend on $S^{|\!\sim}$.

   Thus a database may be a list of the following form.

$$\begin{aligned}\Delta = (&A_1, A_2, \text{ use abduction algorithm } \mathbf{Ab}_1, A_4,\\ &\text{use default algorithm } \mathbf{D}_1, A_6)\end{aligned}$$

3. In fact each item of data in Δ may come with a little *algorithmic patch*, giving the main algorithm $S^{|\!\sim}$ extra or less freedom in using this item of data. A well known example of such a 'patch' is the exclamation mark in linear logic. $!A$ means 'you can use A as many times as you need'. The patch can interact with the *mode* (see section 7 below) and change it to a new mode.

The algorithm $S^{|\sim}$ may approach the third item in the list Δ above with a view of trying to succeed. Instead of a data item it finds an algorithm $\mathbf{Ab}_1$. It exchanges information with $\mathbf{Ab}_1$ and triggers it. What $\mathbf{Ab}_1$ does now depends on the state of the $S^{|\sim}$ algorithm when it 'hits' $\mathbf{Ab}_1$. $\mathbf{Ab}_1$ returns with some additional data, say $(B_1, \ldots, B_k)$. The $S^{|\sim}$ algorithm can now continue *as if* the database were $(A_1, A_2, (B_1, \ldots, B_k), A_4, \mathbf{D}_1, A_6)$. There are two ways of looking at mechanisms. One way is that they are some sort of algorithmic shorthand for additional data, like links to other places where data can be found. Thus according to this view $\Delta +$ *mechanism* $= \Delta + \Delta'$, where $\Delta' =$ data obtained by *mechanism* applied to Δ. This is a traditional view, since a theory Δ is still a (structured) set of data. The second view, which is the one we want to adopt is that procedures are themselves data. This view is better because of two reasons.

1. The mechanisms may yield different results depending on when in the computation they are being used ('hit'), thus making it difficult to say what is the declarative content of the theory.

2. Since we accept the proof theory as part of the logic, we can go all the way and accept additional mechanisms and patches to the proof theory as part of the data. Thus different databases may include as part of themselves additional rules to be used to prove more (or prove less) from themselves. In fact each item of data (declarative unit) can carry as part of its label a patch on the computation $S^{|\sim}$.

This idea is somewhat revolutionary. It also gives up the received view that theories (data) must have a declarative content. It connects nicely with the idea of mode of the next section.

7 MODES OF EVALUATION

When we present logics semantically, through say Kripke models, there are several features involved in the semantics. We can talk about Kripke frames, say of the form (S, R, a), where S is a set of possible worlds, $R \subseteq S^{n+1}$ is the accessibility relation (for an n-place connective $\sharp(q_1, \ldots, q_n)$) and $a \in S$ is the actual world. We allow arbitrary assignments $h(q) \subseteq S$ to atomic variables q and the semantical evaluation for the non-classical connective $\sharp$ is done through some formula $\Psi_\sharp(t, R, Q_1, \ldots, Q_n), Q_i \subseteq S$ in some language. We have:

- $t \vDash \sharp(q_1, \ldots, q_n)$ under assignment h iff $\Psi_\sharp(t, R, h(q_1), \ldots, h(q_n))$ holds in (S, R, a).

For example, for a modality $\Box$ we have

- $t \vDash \Box q$ iff $\forall s(tRs \to s \vDash q)$.

Here $\Psi_\Box(t, h(q)) = \forall s(tRs \to s \in h(q))$. Different logics are characterised by different properties of R.[4]

We can think of $\Psi_\sharp$ as the *mode of evaluation* of $\sharp$. The mode is fixed throughout the evaluation process.

In the new concept of logic, mode shifting during evaluation is common and allows for the definition of many new logics. We can view the mode as the recipe for where to look for the accessible points s needed to evaluate $\Box A$.

Consider the following:

- $t \vDash \Box A$ iff $\forall n \forall s(tR^n s \to s \vDash A)$
  where $xR^n y$ is defined by the clauses:
  - $xR^0 y$ iff $x = y$
  - $xR^{n+1} y$ iff $\exists z(xRz \wedge zR^n y)$.

Clearly $\{(t,s) \mid \exists n tR^n s\}$ is the transitive and reflexive closure of R.

Thus in this evaluation mode, we look for points in the reflexive and transitive closure of R.

We can have several evaluation modes available over the same frame (S, R, a).

Let $\rho_i(x, y, R), i = 1, \ldots, k$, be a family of binary formulas over (S, R, a), defined in some possibly higher-order mode language $\mathcal{M}$, using R, a and h as parameters.

We can have a mode shifting function $\varepsilon : \{1, \ldots, k\} \mapsto \{1, \ldots, k\}$ and let

- $t \vDash_i \Box A$ iff for all s such that $\rho_i(t, s, R)$ holds we have $s \vDash_{\varepsilon(i)} A$.

EXAMPLE 14. Consider now the following definition for $\vDash$ for two modes ρ_0 and ρ_1 and $x = 0$ or 1:

- $t \vDash_x q$ for q atomic iff $t \in h(q)$
- $t \vDash_x \neg A$ iff $t \nvDash_x A$
- $t \vDash_x A \wedge B$ iff $t \vDash_x A$ and $t \vDash_x B$
- $t \vDash_x \Box A$ iff for all s such that $\rho_x(t, s)$ holds we have $s \vDash_{1-x} A$.

We see that we have a change of modes as we evaluate.

We are thus defining $\vDash_0$ not independently on its own but together with $\vDash_1$ in an interactive way.

We repeat here Example 1.5 of [Gabbay, 1998b].

[4] Some logics presented axiomatically, cannot be characterised by properties of R alone but the family of allowed assignments h needs to be restricted. This is a minor detail as far as the question of 'what is a logic' is concerned.

EXAMPLE 15. We consider two modes for a modality $\Box$.

$$\begin{array}{rcl}\rho_1(x,y) & = & xRy \vee x = y \\ \rho_1(x,y) & = & xRy.\end{array}$$

Define a logic $\mathbf{K1}_{[2]}$ as the family of all wffs A such that for all models (S,R,a) and all assignments h we have $a \vDash_0 A$.

In such a logic we have the following tautologies.

$$\begin{array}{l}\vDash \Box A \to A \\ \nvDash \Box(\Box A \to A) \\ \vDash \Box^{2n}(\Box A \to A)\end{array}$$

It is easy to see that this logic cannot be characterised by any class of frames. For let (S,R,a) be a frame in which all tautologies hold and $\Diamond(\neg q \wedge \Box q)$ holds. Then aRa must hold since $\Box A \to A$ is a tautology for all A. Also there must be a point t, such that aRt and $t \vDash \neg q \wedge \Box q$. But now we have $aRaRt$ and this falsifies $\Box^2(\Box A \to A)$ which is also a tautology.

This logic, $\mathbf{K1}_{[2]}$, can be axiomatised. See Example 7.

The idea of a mode of evaluation is not just semantical. If we have proof rules for $\Box$, then there would be a group of rules for logic $\mathbf{L}_1$ (say modal $\mathbf{K}$) and a group for $\mathbf{L}_2$ (say modal $\mathbf{T}$). We can shift modes by alternating which group is available after each use of a $\Box$ rule.

This way we can jointly define a family of consequence relations $\mid\!\sim_\mu$ dependent on a family of modes $\{\mu\}$.

In proof theory the mode can be characterised by what proof rules are available to use in the next step. When a rule is used with an item of data, its 'procedural patch' can change the mode and thus affect the proof process. For example, labelled modus ponens becomes

$$\frac{(\mu_1, s) : A \to B, (\mu_2, t) : A; \varphi(s,t,\mu_1,\mu_2)}{(f(\mu_1,\mu_2), f(s,t)) : B}.$$

So, then, the notion of mode can be attached to data items. Since mode means which proof rules are available, data items can carry mode with them and when they are used they change the mode. This is fully compatible with our approach that procedures are part of the data.

8 TAR-LOGICS (TIME-ACTION-REVISION LOGICS)

We are now ready to introduce TAR-logics. First let us summarise what we so far have ready to hand:

Logical Systems have the following features:

- Structured data; data carrying algorithmic patches;

- algorithmic proof theory on the data structure;
- mechanisms which make use of data and algorithms to extend data;
- a notion of acceptability of a data-structure. This replaces the traditional notion of inconsistency. In fact, inconsistency is no longer a central notion.
- the deduction theorem is connected with cut and the insertion and deletion notions.

Notice the following two points about the current notion of a logical system:

- time, actions and revision are not involved as part of the logic itself;
- proofs and answers are conceptually instantaneous.

Our new notion of logic and consequence shall make the following points:

- proofs take time (real time!);
- proofs involve actions and revisions;
- actions are items of data;
- logics need to be presented as part of a mutually dependent family of logics of various modes.

We can now explain the above points.

In classical geometry, we have axioms and rules. To prove a geometrical theorem may take 10 days and 20 pages, but the time involved is not part of the geometry. We can say conceptually that all theorems follow instantaneously from the axioms.

Let us refer to this situation as a *timeless* situation.

Our notion of a logic developed so far is timeless in this sense. We have a structured database Δ, we have a consequence relation $\mid\!\sim$, we have an algorithm $\mathcal{S}^{\mid\sim}$, we have various mechanisms involved and they all end up giving us answers to the timeless question: does $\Delta \mid\!\sim ?\Gamma$ hold?

In practice, in many applications where logic is used, time and actions are heavily involved. The deduction $\Delta \mid\!\sim ?\Gamma$ is not the central notion in the application, it is only auxiliary and marginal. The time, action, payoffs, planning and strategic considerations are the main notions and the timeless consequence relation is only a servant, a tool that plays a minor part in the overall picture. In such applications we have several databases and queries $\Delta_i \mid\!\sim ?\Gamma_i$ arising in different contexts. The databases involved are ambiguous, multiversion and constantly changing. The users are in constant disagreement about their contents. The logical deductive rules are non-monotonic, commonsense and have nuances that anticipate change and

the reasoning itself heavily involves possible, alternative and hypothetical courses of action in a very real way. The question of whether $\Delta_i \mid\!\sim ?\Gamma_i$ hold plays only a minor part, probably just as a question enabling or justifying a sequence of actions.

It is therefore clear that to make our logics more practical and realistic for such applications, we have no alternative but to bring in these features as serious components in our notion of what is a logic.

Further, to adequately reason and act we need to use a family of different logics at different times, $\Delta_i \mid\!\sim_i ?\Gamma_i$, and their presentation and proof theory are interdependent. We anticipate a heavy use of modes in the deduction.

Let $\{\mathbf{a}_1, \ldots, \mathbf{a}_n\}$ be a family of actions of the form $\mathbf{a}_i = (\alpha_i, \beta_i)$, where α_i is the precondition and β_i is the post condition. Let $*$ be a revision operation such that for a given Δ and α $\Delta * \alpha$ is the result of inputting α into Δ and then revising the new theory into an *acceptable* new theory $\Delta * \alpha$.

We now motivate the idea that actions can be part of the database. We do this in stages. Consider an example.

A lecturer is not performing properly. There are complaints about her/his classes. The Head of Department puts forward the following two very natural statements.

1. We need to issue a formal warning first (W). We cannot just dismiss her/him (D). If his/her behaviour continues (B) then he/she is to be dismissed.

2. If we dismiss him/her, he/she would surely appeal (A).

3. This seems bound to happen. Better call our lawyer now (C).

If Δ is the background theory then in the context Δ the actions of issuing a warning $\mathbf{w}$ requires the precondition B and post condition W.

The action of dismissal $\mathbf{d}$ requires the precondition $W \wedge B$ and then the post condition is D. The action of appeal $\mathbf{a}$ has the precondition D and postcondition A.

The way to formalise the above is as follows:

- B holds now, $\Delta \vdash B$
- We are going to update Δ by issuing a warning and get $\Delta * W$.
- We think B will continue to hold ($\Delta * W \vdash B$).
- We will then dismiss the teacher $(\Delta * W) * D$.
- The teacher will then appeal $((\Delta * W) * D) * A$.
- We therefore need to contact our lawyer now

$$Possible(((\Delta * W) * D) * A) \to C.$$

Clearly available actions should be part of data. Our database Δ comes with actions $\mathbf{w}, \mathbf{d}, \mathbf{a}$. Provability depends on the actions. We can see that a sequence of actions is bound to happen and we are contacting the lawyer now.

We define by induction on n the notion $\Delta \mid\!\sim_{(\mathbf{a}_1,\ldots,\mathbf{a}_n)} A$ as follows:

- $\Delta \mid\!\sim_{\varnothing} A$ iff $\Delta \mid\!\sim A$.
- $\Delta \mid\!\sim_{(\mathbf{a}_1,\ldots,\mathbf{a}_{n+1})} A$ iff $\Delta \mid\!\sim_{\varnothing} \alpha_1$ and $\Delta * \beta_1 \mid\!\sim_{(\mathbf{a}_2,\ldots,\mathbf{a}_{n+1})} A$.

In $\mid\!\sim_{(\mathbf{a}_1,\ldots,\mathbf{a}_n)}$ the sequence $(\mathbf{a}_1, \ldots, \mathbf{a}_n)$ acts as a mode of provability. See Gabbay [2000]

Let us write the deduction theorem as a Ramsey Test:

- $\Delta \mid\!\sim A \to B$ iff $\Delta * A \mid\!\sim B$

Then we can write for an action $\mathbf{a} = (A', A)$:

- $\Delta \mid\!\sim \mathbf{a} \to B$ iff either $\Delta \not\mid\!\sim A'$ or $\Delta * A \mid\!\sim B$.

$*$ is a revision operator. The ordinary deduction theorem can be written as

$$\Delta \vdash (\top, A) \to B \text{ iff } \Delta \cup \{A\} \vdash B.$$

$(A', A) \in \Delta$ means that if $\Delta \mid\!\sim A'$ then Δ can be changed to $\Delta * A$.

There are commercial databases with actions which can be triggered automatically. For example

- if condition A' holds then remove A

or

- if stock level of commodity a' is below 5 then order 50 more items.

9 RELEVANCE, ABDUCTION, QUANTITATIVE REASONING, PROBABILITY, FUZZINESS

Relevance and abduction are a central concepts for the general theory of logical systems. In fact we are writing several books on these and related notions e.g. [Gabbay and Woods, 2001a]. A logical system should be presented as ($\mid\!\sim, S^{\mid\sim}$, *Relevance*, *Abduction*, *Mechanisms*). We would not go into details here but only hint. In non-monotonic logic, the fact that a Δ proves (or does not prove) A can be changed by adding (or deleting) items of data. In a real life argument involving a database, one can argue that more data is 'relevant' and hence by adding the relevant data the provability of A

can change. The additional data may be abduced from some observations. Without a proper notion and algorithmic criteria for relevance and abduction in the logic any A can be proved or not proved by simply extending the data at will.

Formally the abduction and relevance notions can be defined using the algorithmic proof process as follows.

The algorithm for $\Delta \mid\!\sim ?A$ has two kinds of rules:

1. *Reduction rules*
   These have the form

   $$\Delta \mid\!\sim A \text{ if } \bigvee_i \bigwedge_j \Delta_{ij} \mid\!\sim B_{ij}$$

   where the problem of $\Delta_{ij} \mid\!\sim ?B_{ij}$ is *simpler* according to some measure.

2. *Immediate rules*
   These have the form

   2.1. Immediate success: $\Delta \mid\!\sim A$

   2.2. Immediate failure$\Delta \not\mid\!\sim A$

   The immediate success or failure can be recognised immediately by looking at the surface syntactical structure of the problem $\Delta \mid\!\sim ?A$.

The options (1), (2.1) and (2.2) are exhaustive and mutually exclusive.

Any metapredicate such as *Abduce*(Δ, A, X) or *metapredicate*(Δ, A, X) can define X by recursion on the steps (1) and (2). It has to say what the options for X are in the case of immediate success or failure and to say how to construct the options for the Xs of the case $\Delta \mid\!\sim ?A$ if the Xs for the case $\Delta_{ij} \mid\!\sim ?B_{ij}$ are all known.

Of course the recursive procedures must be intuitive and meaningful relative to the intuitive meaning of the metapredicate we want to define.

An additional metalogic may be needed to classify and prioritise the options X arising in the calculation of *metapredicate* (Δ, A, X) for $\Delta \mid\!\sim ?A$.

The notion of relevance of X to the problem $\Delta \mid\!\sim ?A$ can be defined in general as relating or affecting the computation of various metapredicates involving Δ and A. If actions are involved and we need that $\Delta \mid\!\sim A$ to hold so that A will trigger a desired action, then X can be deemed relevant if it can facilitate the abduction of some Xs needed to prove A.

In fact, good definitions can offer a fairly good model of relevance. This subject requires booklength study, but the above describes the basic initial framework.

Quantitative reasoning can be put into the framework via the labelling and via the actions. If our database has labels (i.e. we talk about

$\{t_i : A_i\} \mid\!\sim^{?} s : B)$ then the labels can be multiple labels containing numerical information of any kind. The Dempster–Shafer rule for example has the exact structure of a labelled modus-ponens rule. See [Gabbay, 1996, p. 69].

The presence of actions, especially when they are non-deterministic allows us to bring in concepts and methods from decision theory. Utility measures can be imposed on theories and utility maximisation can be added to the notion of $\Delta \mid\!\sim_{(\mathbf{a}_1,\ldots,\mathbf{a}_n)} A$. All of these avenues need to be further explored and modelled.

10 CONCLUSION

We conclude this chapter with our most recent conception of a logical system for practical reasoning. This idea is yet to be fully explored but it shows the direction in which logic is going. The basic assumption underlying all that was said up to now in this chapter is the *static* question of $\Delta \mid\!\sim ?A$. Δ is given, A is given and we ask "Does A follow"? We made Δ complex, we added mechanisms, metapredicates, actions probability algorithms, but everything was done in the context of there being some given Δ. Even the idea of revision is in this context. We revise Δ for whatever reason to a new Δ' which acts as *the new given database* for the moment.

We now want to break away from this assumption. Chapter one above gave some of the motivation and indicated certain lines of development. The new notion of a logical system is that of (possibly several) perpetual active processes of proof and mechanisms which indefinitely work against an infinite stream of incoming data. (It is like a perpetually working operating system of a big machine.) Different logics are different processes. There is no 'database' as such, just the history of what has come, what was done, and judgements as to what is to be done. The new concept of a logical system is a non-stop logic machine! (An intelligent ever interacting logical agent!)

We leave the reader with this idea.

For more about logical systems and consequence relations see [Gabbay, 1996] as well as [Gabbay, 2000]. For the goal directed computation see [Gabbay, 1998] and [Gabbay and Olivetti, 2000]. For an approach to the dynamics of practical reasoning see [Gabbay and Woods, 2001a]. For an account of actions as premisses see [Gabbay and Woods, 1999b].

BIBLIOGRAPHY

[Alan Ross Anderson, 1975] Nuel D. Belnap Jr. Alan Ross Anderson. *Entailment: The Logic of Relevance and Necessity*, volume 1. Princeton University Press, Princeton, NJ, 1975.

[Aliseda-LLera, 1997] Atocha Aliseda-LLera. *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence.* Amsterdam: Institute for Logic, Language and Computation, 1997.

[Gabbay, 1998] D. M. Gabbay. *Elementary Logic: A Procedural Perspective.* Prentice-Hall, 1998.

[Gabbay, 1998b] D. M. Gabbay. *Fibring Logics* Oxford University Press, 1998.

[Gabbay and Hunter, 1991] D.M. Gabbay and A. Hunter. Making inconsistency respectable part 2: meta-level handling of inconsistency. In LNCS 747, Berlin: Springer-Verlag, 1993.

[Gabbay and Kempson, 1991] D. M. Gabbay and R. Kempson. Natural language content: A proof theoretic perspective. In *Proceedings of the Eighth Amsterdam Colloquium*, pages 173–196, 1991.

[Gabbay and Olivetti, 2000] D. M. Gabbay and N. Olivetti. *Goal Directed Proof Theory.* Kluwer, 2000.

[Gabbay and Reyle, 1984] D. M. Gabbay and U. Reyle. N-prolog, an extension of prolog with hypothetical implication. *Journal of Logic Programming*, 4:319–355, 1984.

[Gabbay and Woods, 1999b] Dov M. Gabbay and John Woods. Cooperate with your logic ancestors. *Journal of Logic, Language and Information*, 8, 1999.

[Gabbay and Woods, 2000a] Dov M. Gabbay and John Woods. Non-cooperation in dialogue logic. *Synthese*, 2000a.

[Gabbay and Woods, 2000b] Dov M. Gabbay and John Woods. More on non-cooperation logic. *Logic Journal of the IGPL*, 9:321–339, 2000.

[Gabbay and Woods, 2001a] Dov M. Gabbay and John Woods. *Agenda Relevance: A Study in Formal Pragmatics.* 2001a. To appear.

[Gabbay and Woods, 2002] Dov M. Gabbay and John Woods. *The Reach of Abduction.* 2002. To appear.

[Gabbay, 1985] D. M. Gabbay. N-prolog, part ii. *Journal of Logic Programming*, 5:251–283, 1985.

[Gabbay, 1990] Dov M. Gabbay. Editorial. *Journal of Logic and Computation*, 1990.

[Gabbay, 1994] Dov M. Gabbay. *What is a Logical System?* Oxford: Oxford University Press, 1994.

[Gabbay, 1996] Dov M. Gabbay. *Labelled Deductive Systems.* Oxford: Oxford University Press, 1996.

[Gabbay, 2000] D. M. Gabbay. What is a logical system II? In *SEP Proceedings, Volume 1*, pp. 81–104, Oxford: Hermes Science, 2000.

[Kaplan, 1996] M. Kaplan. *Decision Theory as Philosophy.* Cambridge: Cambridge University Press, 1996.

[Keynes, 1921] John Maynard Keynes. *A Treatise on Probability.* London: Macmillan, 1921.

[Olivetti and Terracini, 1992a] N. Olivetti and L. Terracini. N-prolog and equivalence of logic programs. *Journal of Logic, Language and Information*, 1:253–340, 1992.

[Pearl, 1988] Judea Pearl. *Probabalistic Reasoning in Intelligent Systems: Networks of Plausible Inference.* San Mateo, CA: Morgan Kaufmann, 1988.

[Pearl, 2000] Judea Pearl. *Causality: Models, Reasoning and Inference.* Cambridge: Cambridge University Press, 2000.

[van Benthem, 1996] Johan van Benthem. *Exploring Logical Dynamics.* Stanford: CSLI Publications, 1996.

[van Benthem, 1999] Johan van Benthem. Resetting the bounds of logic. *European Review of Philosophy*, 4:21–44, 1999.

[van Eemeren and Grootendorst, 1983] van Eemeren and Grootendorst. *Speech Acts in Argumentative Discourse.* Dordrecht: Foris Publication, 1983.

[Woods and Walton, 1972] John Woods and Douglas Walton. On fallacies. *Journal of Critical Analysis*, 5:103–111, 1972. reprinted in *Fallacies: Selected Papers 1972-1982* Dordrecht and Providence, RI: Foris Publications 1989; 1–10.

[Woods and Walton, 1989] John Woods and Douglas Walton. *Fallacies: Selected Papers 1972-1982.* Berlin and New York: Foris de Gruyter, 1989.

[Woods, 1999] John Woods. Hasty generalization. In *Proceedings of the World Congress of Philosophy.* Boston: Philosophy Documentation Center, 1999.

[Woods, 2001c] John Woods. *Death of Argument: Fallacies and Other Distractions*. Newport Beach: Vale Press, 2001.

[Woods, 2001d] John Woods. *Aristotle's Earlier Logic*. Oxford: Hermes Science Publications, 2001.

[Woods, 2001a] John Woods. *Paradox and Paraconsistency: Conflict Resolution in the Abstract Sciences*, Cambridge University Press, Cambridge, 2001. In press.

[Woods, 2002] John Woods. John Locke. *Argumentation*, 2002. To appear.

INDEX